

The University of Chicago

**THE MEASUREMENT AND IMPROVE-
MENT OF SILENT READING AT THE
JUNIOR-HIGH-SCHOOL LEVEL**

**A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF THE
SOCIAL SCIENCES IN CANDIDACY FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY**

DEPARTMENT OF EDUCATION

**By
ARTHUR EDWIN TRAXLER**

**Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS**

1932

The University of Chicago

THE MEASUREMENT AND IMPROVE-
MENT OF SILENT READING AT THE
JUNIOR-HIGH-SCHOOL LEVEL

A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF THE
SOCIAL SCIENCES IN CANDIDACY FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATION

By
ARTHUR EDWIN TRAXLER

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

1932



ACKNOWLEDGMENT

The writer is deeply indebted to the following persons: to Dr. William S. Gray, for advice at all stages of the investigation; to Dr. Charles H. Judd and Dr. Karl J. Holzinger for critical reading of the manuscript; to Dr. Nelson B. Henry for counsel in identifying the problems and launching the study; to Dr. Frank N. Freeman for advice concerning the problem studied in Chapter VI; to Ivan A. Booker, with whom the writer cooperated in studying a number of the basic problems of test construction; to Helen Lacey Iles, who did the teaching in the experiment in corrective reading; to Samuel Scoville, Jr., for permission to reprint the stories used in the rate tests; to numerous principals and teachers who cooperated in the experimentation with the reading-test material; and to his wife, Mabel Yearout Traxler, who rendered much assistance throughout the study.

TABLE OF CONTENTS

	Page
LIST OF TABLES	
Chapter	
I. INTRODUCTION	1
The Purpose of the Study	
Major Steps of Procedure	
Scope of the Investigation	
Organization of the Report	
II. THE ORIGIN AND SIGNIFICANCE OF THE STUDY	4
The Problem of the Chapter	
The University of Chicago High School	
The Research Program in Reading	
The Reading Tests Used	
Critical Evaluation of the Reading-Test Battery	
The Need for a New Battery of Reading Tests	
Corrective and Remedial Reading	
Lack of Diagnostic Techniques	
The General Need for This Kind of Study	
III. THE CONSTRUCTION OF A BATTERY OF SILENT READING TESTS	19
The Problem of the Chapter	
The Criteria of a Good Test	
Construction of the Test of Rate of Reading	
Construction of the Vocabulary Test	
Construction of the Comprehension Test	
The Complete Battery of Tests	
IV. AN EVALUATION OF THE TESTS	36
The Problem of the Chapter	
Ease of Administration	
Ease and Speed of Scoring	
Objectivity of the Tests	
Equivalence of the Forms of the Tests	
Tentative Norms	
Validity of the Tests	
Reliability of the Tests	
The Utility of the Tests	
V. EXPERIMENTAL TEACHING OF CORRECTIVE READING	58
The Problem of the Chapter	
The Initial Experiment	
The Experiment in 1930-31	
Testing Reading Ability	

Chapter	Page
Analysis of Test Scores and Selection of Reading Cases	
Selecting the Control Groups	
Diagnosing the Difficulties	
Planning the Teaching Procedures to Meet the Reading Needs	
The Teaching	
Results of the Experiment	
The Permanence of the Gains in Reading	
Summary of Findings	
VI. THE RELATION BETWEEN RATE OF READING AND SPEED OF ASSOCIATION	80
The Problem of the Chapter	
The Technique Used in Studying the Problem	
Validity and Reliability of the Association Tests	
Correlation of Reading Rate with Speed of Association	
Summary of Facts about Association Time and Rate of Reading	
How the Technique Developed in this Chapter Can Be Used	
VII. CONCLUSIONS AND INTERPRETATION	191
The Measurement of Silent Reading	
Corrective Teaching of Reading	
Relation of Rate of Reading to Speed of Association	
The Need for Continuous Investigation	
BIBLIOGRAPHY	200
APPENDIX	207

LIST OF TABLES

Table	Page
I. Reliability of the Monroe Standardized Silent Reading Test and the Thorndike-McCall Reading Scale . .	7
II. Correlations between Rate Tests and between Comprehension Tests	8
III. Reliability of the Sangren-Woody Reading Test . . .	11
IV. Means and Standard Deviations of the Scores of Seventy-Two Pupils in the University High School on the Four Forms of the Rate Test	24
V. Reliability Coefficients for the Fifty-Word Vocabulary Test	28
VI. Means, Differences between Means, Probable Errors of Differences between Means, and Standard Deviations of Scores Made by Three Groups of Pupils on Various Forms of the Rate Test	38
VII. Means, Differences between Means, Probable Errors of Differences between Means, and Standard Deviations of Scores Made by 208 Ninth-Grade Pupils on the Four Forms of the Vocabulary Test	39
VIII. Means, Differences between Means, and Probable Errors of Differences between Means of Scores Made by Pupils at the Junior-High-School Level on the Various Forms of Part I and Part II of the Comprehension Test	40
IX. Means, Differences between Means, Probable Errors of Differences between Means, and Standard Deviations of Scores Made by Three Groups of Pupils on Various Forms of the Combined Comprehension Score .	42
X. Means, Differences between Means, Probable Errors of Differences between Means, and Standard Deviations of the Scores Made by Three Groups of Pupils on Various Forms of the Entire Test Battery	43
XI. Mean Scores Made by the Pupils in the Lowther Junior High School, Emporia, Kansas, on Form 1 of the Silent Reading Test which was Constructed in This Study	44
XII. Correlations between the Constructed Vocabulary Test and the Inglis Test of English Vocabulary . .	46
XIII. Correlations Among Various Comprehension Tests . .	48

Table

Page

XIV. Correlations of Test Scores with Average School Marks	51
XV. The Reliability of the Test of Rate of Reading . .	53
XVI. Reliability of the Vocabulary Test	54
XVII. Reliability of the Total Comprehension Score . . .	54
XVIII. Reliability of the Composite Score	55
XIX. The Reliability of the Reading Tests When Several Forms are Used Consecutively	56
XX. Gains in Reading Made by the Experimental Group and the Control Group in the Preliminary Experiment in the Seventh Grade	60
XXI. Scores of Corrective Group on Entrance Tests in Reading	64
XXII. Mean Scores Made by the Pupils in the Corrective Reading Class and in the Control Group on the Various Parts of the Sangren-Woody Reading Test	72
XXIII. Mean Scores Made by the Experimental Group and the Control Group on Forms One, Two, and Three of the U. H. S. Reading Test, Administered on February 10, April 21, and June 12, Respectively	73
XXIV. Mean Gains in Test Scores Made by the Reading Class	74
XXV. Number and Percentage of Pupils in Experimental Group below Norm when Sangren-Woody, Gray, and U. H. S. Tests were First Administered Compared with Number and Percentage below Norm when the Tests were Given at the End of the Year	75
XXVI. The Gains in Mean Scores Made on the U. H. S. Test by the Pupils in the Experimental and Control Groups during the Summer Vacation following the Experiment	78
XXVII. The Reliability of the 46-Word Forms of the Free-Association Test and the 100-Word Forms of the Test of Antonyms	88
XXVIII. The Mean Number of Seconds Required to Perform the Association Tests	88
XXIX. Correlation of Speed of Association with Rate of Reading	90
XXX. Correlation of Association Time and Rate of Reading with Intelligence	91
XXXI. Partial Correlation of Speed of Association with Rate of Reading when Intelligence is Held Constant	92

Table	Page
XXXII. Correlation of Association Time and Rate of Reading with Vocabulary	94
XXXIII. Partial Correlation of Speed of Association with Rate of Reading when Vocabulary is Held Constant .	95

CHAPTER I INTRODUCTION

The Purpose of the Study

Research studies carried on in recent years have shown clearly that, notwithstanding improved methods of teaching reading, many pupils pass through the first six grades of the elementary school without acquiring the skills needed to enable them to do effectively the extensive silent reading which the higher levels of the school demand.¹ It has been found further that there is a significant positive relationship between ability to read silently and scholastic success in the junior and senior high schools and in college.² Faced with these facts, administrators and teachers in the upper grades and the junior high school are becoming

¹I. Jewell Simpson and Bessie C. Stern, "Improving Instruction in Silent Reading," Elementary School Journal, XXV (April, 1925), 594-606.

A. H. Sutherland, "Correcting School Disabilities in Reading," Elementary School Journal, XXIII (September, 1922), 37.

William S. Gray, "Case Studies of Reading Deficiencies in Junior High School," Journal of Educational Research, X (1924), 132-33.

P. B. Jacobson and E. C. Vandusen, "Remedial Instruction in Reading in the Ninth Grade," School Review, XXXVIII (February, 1930), 142-43.

Pearl Sands McCarty, "Increasing Comprehension in Silent Reading," School Review, XXXIX (December, 1931), 758.

²William S. Gray, op. cit., 133.

Unpublished Report of the Remedial Workers in Reading in the University High School, 1929-30; filed with the Research Secretary of the Laboratory Schools, the University of Chicago, Chicago, Illinois.

Elbert K. Fretwell, A Study of Educational Prognosis. Teachers College Contributions to Education, No. 99. New York: Teachers College, Columbia University, 1919.

Bertha M. Smith, "Correlation of Ability in Reading with the General Grades in High School," School Review, XXVII (September, 1919), 493-511.

increasingly aware of the necessity of adopting measures to correct the reading defects of a large percentage of their pupils and thus to remove handicaps which are serious barriers to school progress. In response to this need, more than a score of experiments in teaching remedial and corrective reading to groups of retarded readers have been conducted within the last decade.¹ Although a number of these investigations have made valuable contributions to the chain of evidence in this field, the results of the experimental work are, nevertheless, tentative and inconclusive. The development of adequate techniques for the improvement of the silent reading ability of pupils above the elementary school level has awaited the findings of additional research studies.

The purpose of this investigation is to study statistically and experimentally a series of problems relating to the measurement and improvement of silent reading achievement at the junior-high-school level. The following closely related problems are involved: (1) the development of improved means of measuring silent reading achievement; (2) the development of corrective measures that are economical of time and that can be used by classroom teachers; and (3) the invention of special diagnostic techniques that will supplement those that are now available for use in determining the specific nature of reading deficiencies. The origin and significance of each of these problems will be outlined in Chapter II.

Major Steps of Procedure

There are three major steps in this study, corresponding to the three problems that have been stated. The nature of the steps is indicated in a general way in the following statements:

1. A battery of silent reading tests in four forms, measuring rate of reading, reading vocabulary, and comprehension, was devised and partially standardized.

2. An experiment in teaching silent reading to a group of retarded readers at the junior-high-school level was conducted. In this experiment, corrective procedures applicable to a classroom situation were devised, the effectiveness of the procedures in improving the reading ability of the pupils was evaluated, and the utility of the tests constructed in the first part of the study was demonstrated.

¹William S. Gray, Summary of Investigations Relating to Reading, Supplementary Educational Monographs, No. 28, Chicago, Department of Education, University of Chicago, 1925, and Supplements 1-6.

3. Various needed diagnostic techniques for making a complete analysis of reading deficiencies were identified during the experimental teaching and one of them was developed. The new technique consists of a procedure for studying the relationship between rate of reading and rate of association.

The methods used in each of the steps will be explained in detail in subsequent sections relating to the various parts of the study.

Scope of the Investigation

In the construction of the tests, use was made of the scores of more than twenty-five hundred pupils in Grades VII to IX, distributed among the University of Chicago High School and nine public schools.

Use was made of several groups of pupils in the University High School¹ in the reading experiment. A class of twenty-seven seventh-grade pupils comprised the experimental group in the introductory study, and another seventh-grade class acted as a control group. In the main study, a class of twenty-four seventh-grade pupils formed the experimental group, and two groups, each consisting of twenty-nine pupils, served as control groups.

In the study of a new diagnostic technique in reading, the scores of twenty-seven Freshmen in the University of Chicago and of ninety seventh-grade pupils and sixty-seven ninth-grade pupils in the University High School were used.

Organization of the Report

Following Chapter II, in which the origin and significance of the study are outlined, the organization of the report conforms to the major steps of procedure that were listed in the second section of this chapter. The construction and standardization of improved instruments for the measurement of silent reading are considered in Chapters III and IV; the experimental study of the use of the new tests and the application of corrective measures to the reading problems of a group of junior high-school pupils is presented in Chapter V; the development of a supplementary diagnostic technique and its use in corrective work with retarded readers is described in Chapter VI. The investigation is summarized and interpreted in Chapter VII.

¹In this study, the words "University High School" will always mean the University of Chicago High School.

CHAPTER II

THE ORIGIN AND SIGNIFICANCE OF THE STUDY

The Problem of the Chapter

The purpose of this chapter is to discuss the origin and significance of the specific studies involved in this investigation. In order to describe clearly the conditions and needs which gave rise to it, the program of remedial and corrective reading in the University High School will be explained briefly.

The University of Chicago High School

The University High School is a part of the University Laboratory Schools which provide eleven years of school training above the kindergarten and which lead to a high school diploma equivalent to those customarily granted at the end of a twelve-year program. The first six years of work are in the University Elementary School; the last five years are in the University High School. The high-school years are known as the Sub-Freshman, Freshman, Sophomore, Junior, and Senior years. Thus, the Sub-Freshman year is the seventh grade and the Freshman year is the eighth grade for the pupils who come from the University Elementary School. However, more than 50 per cent of the pupils who enter the Freshman year are graduates of other elementary schools and are admitted after they have completed eight years of school work. The Freshman year is, therefore, the ninth grade for them. In order to make the terminology in this study agree as closely as possible with that which is in general use, the first or Sub-Freshman year of the high school will be called the seventh grade and the Freshman year will be known as the ninth grade.

The Research Program in Reading

In addition to the studies that are carried on by the regular staff of teachers, a program of research has been fostered in the University High School for several years through either a full-time research assistant or two half-time assistants. Although their functions are quite broad in scope, the major portion of the time and energy of the research workers has been directed to reading problems because the most pressing need for research has been in that field.

A battery of tests has usually been administered by the

research workers to all pupils at the time of their entrance to the University High School. The pupils who come from the University Elementary School take the tests when they enter the seventh grade; the pupils who are received from other elementary schools take them when they enter the ninth grade. The battery of tests has consisted in the past of an intelligence test, several reading tests, an arithmetic test, and a group of short civictests. Various uses have been made of the test scores by members of the high-school staff. The most extensive use of them has been in the discovery of pupils who are retarded in reading and in the diagnosis of their difficulties. Following the diagnosis, a program of remedial teaching has been developed in accordance with the needs of the pupils.

The Reading Tests Used

The majority of the twenty-one reading and vocabulary tests that are suitable for use in Grades VII to IX have been tried out, at various times, with University High School pupils. This process of empirical trial and selection resulted in the choice of the following reading tests for use in the entrance battery: the Gray Silent Reading Test, which measures rate of reading; the Monroe Standardized Silent Reading Test III, which measures rate and comprehension; the Thorndike-McCall Reading Scale, which measures comprehension; and the Inglis Test of English Vocabulary.

It will be observed that the preceding series of tests gives two scores on rate of reading, two on comprehension, and one on vocabulary for each individual. The practice of using these types of tests is based on the cumulative experience of the remedial workers in the University High School. McCallister found in his intensive case work with handicapped readers that, in addition to a mental test, he needed to use tests of rate of reading, reading comprehension, and vocabulary.¹ The remedial workers in reading, 1929-31, likewise found that those kinds of tests were basic to satisfactory diagnosis of inferior reading ability; all three types were used with each case that was studied by them.²

¹J. M. McCallister, "An Analytical Study of Reading Deficiencies of Junior High School Pupils," Chapters III and IV. Unpublished Doctor's Dissertation, Department of Education, University of Chicago, Chicago, Illinois, 1929.

²Unpublished Reports of the Remedial Workers in Reading in the University High School, 1929-31; filed with the Research Secretary of the Laboratory Schools, University of Chicago, Chicago, Illinois.

The experience of these research workers indicates that a thorough analysis of reading deficiencies should not be attempted until all of these aspects of reading achievement have been measured.

The other tests in the battery were the Otis Self-Administering Test of Mental Ability, the Reavis-Breslich Arithmetic Test, the Hill Civic Information Test, the Hill Civic Attitudes Test, and the Hill-Wilson Civic Action Test. The Otis Self-Administering Test of Mental Ability has marked value for supplementing the results of the reading tests, as it has been found from reports made by teachers in the University High School that there is decided correspondence between failure to do satisfactory reading in the content subjects and low scores on the Otis test. The other tests have some value in the selection of inferior readers, but their use in that respect is naturally quite limited, as they are designed and used primarily for other purposes.

Critical Evaluation of the Reading-Test Battery

The part of the battery of entrance tests which deals with reading has never been thought by the research workers to be completely satisfactory; rather, it has been regarded as the best available makeshift for getting several independent measurements of the reading ability of the new pupils. The most important weaknesses of the battery of tests may be summarized as follows: (1) shortcomings revealed in the scores when they are treated statistically; (2) lack of sufficient number of duplicate forms of some of the tests; (3) time consumed in the administration of the tests; (4) the fact that no one test measured several aspects of reading ability.

Statistical shortcomings.- Two of the four reading tests that are in the entrance battery appear to be rather low in reliability. Table I shows reliability coefficients for the Monroe Standardized Silent Reading Test and the Thorndike-McCall Reading Scale.¹

The correlation coefficients in Table I were found by correlating scores on alternate forms of the tests that had been administered to pupils in a single grade. As they are based on small numbers of cases, they are not very trustworthy. Nevertheless, as none of the correlations are above .75 and two of them are below .50, the evidence indicates that the tests have but little value for the diagnosis of the difficulties of individual pupils. It

¹All correlation coefficients in this study are computed by the product-moment method, unless otherwise stated.

TABLE I

RELIABILITY OF THE MONROE STANDARDIZED SILENT READING TEST
AND THE THORNDIKE-McCALL READING SCALE

Test	Number of Cases	Reliability	
		r	P.E.
Monroe Rate of Reading	52	.606 ±	.058
Monroe Rate of Reading	26	.492 ±	.101
Monroe Rate of Reading	26	.730 ±	.062
Monroe Comprehension	52	.517 ±	.070
Monroe Comprehension	26	.624 ±	.082
Monroe Comprehension	26	.708 ±	.068
Monroe Comprehension	61	.646 ±	.051
Thorndike-McCall	25	.750 ±	.059
Thorndike-McCall	22	.433 ±	.121

is generally held that the reliability of a test which is to be used in diagnostic work should be at least .80; tests whose reliability falls below that point are not of much value in studying individual pupils.¹

There is some evidence that the reliability of the rate score on the Gray Silent Reading Test is somewhat higher than the rate score on the Monroe Standardized Silent Reading Test, but there are not enough data on that point at hand to warrant a generalization. The Inglis Test of English Vocabulary has a reliability coefficient of .90 or above, which is much higher than the reliability of the other tests.

Not only is the reliability of some of the reading tests in the entrance battery rather low, but, furthermore, the agreement between the tests that are designed to measure the same aspect of reading ability is not very close; that is, the two rate tests do not show very high correlation with each other; neither do the two comprehension tests. This fact is brought out clearly in Table II.

The correlation coefficients in Table II indicate that one of two conditions exists: either the tests that are designed to measure the same thing measure somewhat different things or else they measure the same thing but do not measure it consistently. Either condition seriously limits the value of the results that

¹G. M. Ruch, The Objective or New Type Examination, p. 434. Chicago: Scott, Foresman and Company, 1929.

are secured with the tests.

TABLE II*

CORRELATIONS BETWEEN RATE TESTS AND BETWEEN COMPREHENSION TESTS

Tests Correlated	Number of Cases	<u>r</u>	P.E.
Gray Rate and Monroe Rate	80	.521 $\pm$	.055
Monroe Compr. and Thorndike-McCall . .	80	.580 $\pm$	.050
Monroe Compr. and Thorndike-McCall . .	58	.344 $\pm$	.078
Monroe Compr. and Thorndike-McCall . .	75	.547 $\pm$	.055

*Based on scores made by entering pupils in the University High School.

Number of duplicate forms.- Experience in teaching reading in the University High School indicates that the remedial or corrective teacher should use the same battery of tests at least four times with each reading group. The occasions for its use are as follows: (1) the teacher should administer the battery before instruction is begun in order to locate the poor readers and to diagnose their weaknesses; (2) he should give it again after training has been going on for a few weeks in order to study the effectiveness of his teaching and to find out what modifications in procedure should be adopted; (3) he should use it at the end of the training period in order to find out how much gain the group as a whole has made and to discover the pupils who will probably need "follow-up" teaching; (4) finally, it is highly desirable that he administer the battery of tests some months after training has ended so that he may find out whether or not the corrective or remedial teaching has made permanent changes in the reading habits of the pupils.

Unless the same test is repeated, such a program of testing calls for four forms of the battery of tests. Only a part of the battery of tests that has been used in the University High School meets this standard. The Inglis Test of English Vocabulary contains four forms which are for all practical purposes equivalent to one another. The Gray Silent Reading Test and the Monroe Standardized Silent Reading Test, however, have but two forms, each of which must be used twice in the testing program outlined. The Thorndike-McCall Reading Scale has ten forms, the largest number contained in any reading test, but there is marked overlapping in

the forms, both as to content and to questions on the paragraphs. Numerous opportunities are presented, therefore, for specific practice to influence the results. Thus it appears that the Inglis test is the only part of the battery that is completely satisfactory for purposes of repeated measurement during an extended period of instruction in reading.

When one examines other standardized reading tests that are suitable for use in the junior high school, it is soon discovered that a comprehensive battery of reading tests containing four forms does not exist. The publishers of the New Stanford Achievement Test report that it is to have five forms, but some of them have not yet been published.¹ The Thorndike Test of Word Knowledge has four forms; the Burgess Silent Reading Test likewise has four forms, but it is primarily a test of rate of reading and is used so extensively in the University Elementary School that many of the high school pupils are quite familiar with it. No other reading test that is usable in the junior high school has more than three forms and few of the tests have more than two forms. It is apparent, then, that this shortcoming in the entrance test battery of the University High School could not be removed by replacing the present battery with other standardized tests that are now available.

Time required to administer the test battery.- Including the time needed for the distribution and collection of papers and the reading of directions, the amount of time required to administer the reading test battery that has been in use in the University High School is approximately as follows:

Monroe Standardized Silent Reading Test	10 minutes
Gray Silent Reading Test	15 minutes
Thorndike-McCall Reading Scale	35 minutes
Inglis Test of English Vocabulary	45 minutes
Total Time	105 minutes

At least two full class periods are needed for the reading tests. The fact that the battery must be administered in two sittings, if it is given during the regular school day, complicates the use of it. Furthermore, if it is administered several times during a year, in order to measure progress at intervals and thus to direct the teaching to better advantage, an excessive amount of class time is used up in this way. This is especially objectionable, if a control group is used as a check on the progress

¹All the forms of the Stanford Achievement Test are now available. The fifth one appeared after the completion of this study.

of the experimental group, as such extensive testing seriously interrupts the continuity of work in the control class.

Number of aspects of reading ability measured by the tests.- One reason why so much time was used in making a comprehensive measurement of reading achievement was that three of the four tests measured only one aspect of silent reading ability, and the fourth measured but two. Thus, if a number of phases of reading ability were to be measured, it was necessary to use several tests, and the mere routine of distributing and collecting test booklets and filling information blanks on each one consumed a considerable portion of a class period. This unsatisfactory condition could be avoided only by securing a different battery of tests.

The Need for a New Battery of Reading Tests

When the testing program in reading in the University High School was studied in the light of the preceding facts, it was apparent to the research workers that a change in the battery of reading tests was desirable. An attempt was first made to find a more suitable test battery among the standardized tests that had not yet been tried out in the University High School. The newer reading tests and the literature about such tests were canvassed with considerable care. This survey led to the conclusion that, from the standpoint of the needs of the reading program in the University High School, the Sangren-Woody Reading Test, published in 1927, contained the most promising battery.¹

This test contains the following parts: word meaning, rate of reading, facts, central thought, total meaning, following directions, and organization. It has one significant shortcoming due to the fact that it contains only two forms, but, on the other hand, it has two points in its favor: it can be administered in about thirty minutes and it is reported to have very high reliability. Sangren reports for the various parts of the test fourteen reliability coefficients which average .87 and two reliability coefficients for the composite score on the battery which average approximately .96.² Thus, the Sangren-Woody test seemed to meet two of the objections that had been raised against the reading tests that were being used in the University High School, namely, time required to administer them and lack of consistency in the

¹Paul V. Sangren and Clifford Woody, The Sangren-Woody Reading Test. Yonkers-on-Hudson: The World Book Company, 1927.

²Paul V. Sangren, The Measurement of Silent Reading, p. 32. Kalamazoo, Michigan, 1927.

scores.

Accordingly, the Sangren-Woody Reading Test was tried out with eighty-four pupils in the seventh grade of the University High School. Form A was administered on Friday and Form B was given on the following Monday. This procedure corresponded to the technique that was used by Sangren in getting the scores on which he based his reliability coefficients. The correlations between the two forms of the test are shown in Table III.

TABLE III

RELIABILITY OF THE SANGREN-WOODY READING TEST

Part	Number of Cases	<u>r</u>	P.E.
Word Meaning	84	.677 $\pm$	.040
Rate of Reading	84	.544 $\pm$	.051
Fact Material	84	.697 $\pm$	.038
Total Meaning	84	.770 $\pm$	.029
Central Thought	84	.703 $\pm$	.037
Following Directions	84	.646 $\pm$	.043
Organization	84	.456 $\pm$	.058
Average Coefficient		.642	
Total Score	84	.839 $\pm$	.022

The coefficients of correlation in this table are, in all cases, decidedly lower than the ones that were reported by Sangren. They indicate that, for small groups such as were tested in this case, the reliability of the parts of the Sangren-Woody Reading Test is not very high. The reliability of the total score is fairly satisfactory, but it is not exceptionally high. It appears that the test may be used for purposes of making group comparisons, but that the reliability of the various parts is scarcely high enough to make the test of much value in individual diagnostic work.

It was decided, therefore, that, whereas the Sangren-Woody Reading Test could be used to supplement the battery of tests already in use in the University High School, the ultimate aim of the research workers could not be realized until a reading test that was more reliable when used with small groups had been developed. In fact, it was evident that the development of a satisfactory program of remedial and corrective work in reading was dependent to a very considerable degree upon the use of measuring

instruments that were both more serviceable and more reliable.

Stated concisely, the need was for a battery of reading tests which would meet the following conditions:

1. It should measure rate of reading, reading vocabulary, and comprehension, since the fact had been brought out in all case work with retarded readers in the University High School that effective diagnosis invariably involves a knowledge of these three aspects of reading ability.¹

2. It should conform to the generally accepted criteria of a good test, such as high validity, reliability, and objectivity.

3. For reasons already pointed out, it should have four equivalent forms.

4. In order to have a high degree of usefulness for classroom teachers, it should be designed for administration within one class period and for rapid and accurate scoring.

As no battery of tests met all of these conditions, it became, out of sheer necessity, one of the major objectives of this study to devise such a battery. The construction and evaluation of the tests will be reported in the next two chapters.

Corrective and Remedial Reading

The origin of the second part of the study will be explained in this section.

Definition of terms.- Corrective reading is understood to apply to the procedure that is used with pupils who are retarded readers, but whose difficulties will, nevertheless, respond to treatment within a group through the use of methods commonly employed in classrooms. Remedial reading, on the other hand, designates the techniques adopted with those retarded pupils who are so seriously handicapped in the basic reading skills that they require the use of unusual procedures and techniques, and individual attention and guidance in overcoming their deficiencies.

Brief description of the individual technique.- During the last decade, techniques for teaching remedial reading to individual pupils were worked out quite thoroughly in the University Laboratory Schools.² Individual techniques were used in all cases both

¹Unpublished reports of the Remedial Workers in the University High School for the years 1927-28, 1928-29, 1929-30, and 1930-31. These reports are filed in the office of the Research Secretary of the Laboratory Schools, University of Chicago, Chicago, Illinois.

²William S. Gray with the cooperation of Delia Kibbe, Laura Lucas, and Lawrence W. Miller, "Remedial Cases in Reading:

corrective and remedial. These techniques consisted essentially in the application of the methods and procedures of case study to reading difficulties. When tests, observation, or teachers' reports revealed that a child was seriously retarded in reading, the remedial worker undertook a careful study of the case. He began by collecting every available item of information concerning the pupil. He looked up the child's in-school and out-of-school history; he went over his physical record; he talked with the child and had conferences with the child's parents; he gave the pupil other reading tests which were designed to measure particular skills; he often made an intensive study of the pupil's reading habits through the use of the photographic technique. When the data had been assembled, the remedial worker analyzed them, diagnosed the difficulty as best he could, and outlined a course of remedial treatment.

In carrying out this program of remedial treatment, the case worker scheduled meetings with the pupil either at his study periods or after school hours. The training was directed toward certain definite ends, a careful record of progress was kept, and the case was discharged as soon as test scores and class work indicated that the individual had reached a satisfactory level of achievement in reading.

Although, in theory, the remedial teaching was individual in character, in practice it was not always carried on as such. Sometimes several pupils manifested the same reading difficulties. These cases were handled in a small group, provided the time schedule could be arranged to meet the convenience of all of them. Even when the problems were not quite the same for the various individuals, several of them frequently worked together on certain phases of reading.

Advantages of the individual plan.- Individual remedial reading has certain definite points in its favor, some of the more important of which are the following:

1. It has back of it a precedent of long standing. Although the case-study technique, as such, is new, the practice of giving "slow" pupils special help to enable them to "catch up" with their fellows is very old. Since individual remedial teaching

Their Diagnosis and Treatment," Supplementary Educational Monograph No. 22. Chicago: Department of Education, University of Chicago, 1922; James M. McCallister, "An Analytical Study of Reading Deficiencies of Junior High School Pupils." Unpublished Doctor's Dissertation, School of Education, University of Chicago, 1929.

follows a familiar, accepted way of doing things, it tends to inspire confidence in pupil and parent alike.

2. Individual remedial teaching allows the case worker to become well acquainted with the pupil and to study his reading weaknesses at close range.

3. In individual remedial teaching the training is concentrated upon those points where the greatest weakness is shown. The remedial worker, after he has made his diagnosis and has set up the objectives to be attained, can drive toward those objectives, unhampered by any necessity of taking the needs of other pupils into consideration.

4. A few pupils experience such marked difficulty in acquiring the fundamental reading habits that they require the kind of careful individual guidance that can be given only by one who is thoroughly familiar with their case history.

Limitations of the individual plan.- The advantages that are inherent in the individual plan of teaching backward readers are offset by a number of limitations. The most serious of these are probably the following:

1. The pupil may acquire an unfavorable attitude toward the reading practice, since it is not a part of his regular class work. As most of his classmates do not have special work in reading, he tends to feel that it is relatively unimportant. He lacks the stimulating influence of competition with others. The reports of the remedial worker of the University High School for twenty-three pupils who received remedial instruction in 1927-28 and 1928-29 show that twelve pupils, or 52.2 per cent, cooperated well in the training; four, or 17.4 per cent, showed fair interest; and seven, or 30.4 per cent, exhibited lack of interest and did not give the desired amount of cooperation to the remedial worker.¹ Although there is no way of determining whether the interest and cooperation would have been better if the pupils had been taught under regular classroom conditions, the percentage of pupils who did not take an interest in the work was larger than the percentage that was found in an experiment which will be described later in which a group of twenty-four pupils was taught corrective reading. In that experiment, only two pupils, or 8.3 per cent, failed to give full cooperation.

2. If study periods are used, the remedial work takes time

¹Unpublished reports of the Research Worker in Remedial Problems for 1927-28 and 1928-29, University High School, University of Chicago, Chicago, Illinois. Filed in the office of the Research Secretary of the Laboratory Schools.

that could profitably be given to preparation for regular class activities. Most poor readers are somewhat backward in all phases of school work and, consequently, need all the time that they can find for studying. In fact, it has been found in the remedial work in the University High School that one of the chief obstacles in the way of teaching remedial reading to pupils during their study periods is that their regular teachers often summon them at those times for special work in the content subjects.

3. At best, meetings between the remedial worker and individual pupils are held infrequently. The number of meetings is conditioned by the number of study periods which the pupil has and by the amount of time that the remedial worker can give to each pupil. During the years 1927-28 and 1928-29, the meetings between the remedial worker and twenty-two retarded readers averaged 1.7 class periods a week for each pupil.¹ Where it was possible to meet the pupils in small groups, the remedial worker was usually able to arrange for two class periods each week, but in the cases where the time schedule or other factors made it necessary for the remedial teacher to work with single individuals, the number of meetings was usually reduced to one each week.

It is obviously somewhat difficult to maintain interest and secure progress under such conditions. This limitation was recognized by the remedial teacher in the following observation on the training of one of the pupils:

"The amount of training (one period a week) was inadequate for a pupil with such serious deficiencies. However, his program of instruction in other courses was such that only one period could be provided."²

4. By its very nature, the technique of meticulous case analysis and individual remedial treatment works a great drain upon the research worker's time. This is, of course, necessary with pupils who have not yet made the fundamental reading adaptation, but it is not defensible with pupils whose difficulties are really corrective in character.

5. In schools which have large enrolments, individual attention to reading difficulties is not often feasible, because it would require a very large staff of workers if every backward reader were to receive effective individual teaching. This becomes evident when one considers the number of backward readers who were taught by the remedial worker in the University High School. In

¹Reports of Research Worker, op. cit.

²Ibid. (Page references cannot be given as the pages of the report are not numbered.)

1927-28 and 1928-29, the school had the services of a full-time remedial worker who gave the major part of his attention to remedial problems in reading. For each reading case that he handled, he made a very thorough diagnosis, prescribed treatment, gave intensive remedial teaching, and measured the results of the training with great care. He trained fourteen pupils in reading in 1927-28 and nine pupils in 1928-29. It may be concluded from these figures that a remedial teacher who devoted all of his time to reading could probably train twenty to thirty pupils thoroughly by the individual method each year.

It is not uncommon to find that 10 to 25 per cent of the pupils in the junior-high-school grades are seriously retarded in reading. Consequently, in some large junior high schools hundreds of pupils need the attention of a specialist in reading. It is evident that if each pupil must be taught individually, very few schools can provide the needed training.

Group corrective reading.- For some years preceding the present study, the research workers in the University High School have been aware of the desirability of introducing group corrective reading on an experimental basis. The reasons for favoring the initiation of group teaching were these:

1. It was recognized that the limitations of the individual technique constituted serious obstacles to the carrying out of a well-rounded program which would reach all backward readers.

2. The results reported in educational journals concerning attempts at corrective reading in other schools suggested that this type of teaching was feasible.¹

3. It was thought that retarded readers might take greater interest in improving their reading if the work were made an integral part of one of their regular courses.

4. Many high schools whose large enrolments precluded individual teaching were trying to initiate programs of corrective reading.² These schools were in need of techniques for handling

¹The articles that report previous experiments in corrective reading are listed in the bibliography. They are not reviewed in this study, as they have already been summarized by William S. Gray in Summary of Investigations Relating to Reading, Supplementary Educational Monographs, No. 28, Chicago, Department of Education, University of Chicago, 1925, and in Supplements 1-6.

²Of eight public high schools in which students were used in trying out the reading tests constructed in this study, four high schools had recently inaugurated, or were about to start, programs of remedial or corrective reading. Two of them enrolled more

their backward readers in groups. It was felt that since the University High School is an experimental school, it should take a leading part in the development of such techniques.

Accordingly, controlled experiments in the teaching of corrective reading to groups of retarded readers in the seventh grade were carried on during parts of the school years 1929-30 and 1930-31. The major purposes of the experiments were:

1. To study experimentally the value of teaching corrective reading to a group of seventh-grade pupils.
2. To develop techniques for use in correcting the reading defects of seventh-grade pupils.
3. To investigate the permanence of gains made during the teaching of corrective reading.

A detailed account of the plans, procedure, progress, and results of this part of the study will be given in Chapter V.

Lack of Diagnostic Techniques

During the course of the training in reading, the corrective teachers found themselves handicapped, at times, by a dearth of appropriate diagnostic techniques. Despite the great progress that has been made in the study of reading habits within the last quarter of a century, especially in the recording of eye-movements, the complete and accurate diagnosis of many types of reading difficulties encountered by individual pupils is still in a very crude stage. Quick and accurate diagnosis will not become possible until additional techniques are developed to meet the needs which arise.

It was outside the scope of this study to attempt to develop laboratory techniques to solve all of the questions that were encountered during the corrective teaching. However, one question—that of the relation of rate of reading to rate of association—was very closely related to teaching procedure with certain slow readers. A special study of it was, therefore, made and practical applications of the technique that was developed were shown. This study will be reported in Chapter VI to illustrate the kind of diagnostic study that must be carried on if corrective work in reading is to be placed on an adequate scientific basis.

The General Need for This Kind of Study

The preceding sections of the chapter have shown how this

than two thousand pupils. The great problem was how to deal effectively with the many retarded readers that were inevitably found in so large a student body.

study originated in the needs of a particular reading program—that of the University of Chicago High School. However, it is evident from the large number of objective studies of the teaching of reading¹ that have been carried on in the junior high school in the last few years that the reading needs of the University High School are not peculiar and that there is wide-spread interest in the improvement of silent reading. The present investigation, then, is in direct line with the fundamental need for scientific studies of the problem of improving reading at the junior high-school level.

¹See W. S. Gray's Summary of Investigations Relating to Reading, op. cit.

CHAPTER III

THE CONSTRUCTION OF A BATTERY OF SILENT READING TESTS

The Problem of the Chapter

The first major problem of this study was the construction of a test battery consisting of four equivalent forms of a rate of reading test, a reading vocabulary test, and a comprehension test suitable for use in the junior high school. The need for such a battery of silent reading tests was shown in Chapter II. The present chapter will report the procedures used in constructing the tests.

The Criteria of a Good Test

An attempt was made to devise tests that were in harmony with certain generally accepted criteria of a good test. The following criteria are the result of an analytical study of several recent books on test construction.¹

1. Ease of administering and scoring.
2. Objectivity, or freedom from the influence of personal opinion in scoring.
3. Availability of duplicate forms.
4. Tentative norms.
5. Validity, or the degree to which a test measures what it is intended to measure.
6. Reliability, or the degree to which a test measures consistently whatever it does measure.
7. Utility.

These criteria were kept constantly in mind during the work of constructing the tests. In order to simplify the presentation in this chapter, however, the application of the criteria will not be introduced into the description of test construction, except at points where some reference to them is necessary in order to

¹G. M. Ruch, The Objective or New-Type Examination. Chicago: Scott, Foresman and Company, 1929;
C. W. Odell, Traditional Examinations and New-Type Tests. New York: The Century Company, 1928;

Harry A. Greene and Albert N. Jorgensen, The Use and Interpretation of Educational Tests. New York: Longmans, Green and Company, 1929.

explain the steps in making the tests. The use of the criteria in the evaluation of the tests and the data upon which the evaluation was based will be considered in detail in Chapter IV.

Construction of the Test of Rate of Reading

The type of rate test to devise.- The various types of standardized tests of reading rate were studied carefully for points of strength and weakness. On the basis of this analysis a decision was reached to construct a rate test which would consist of continuous discourse and in which no attempt would be made to measure comprehension until the subject had read all the material. This kind of test uses reading conditions under which children ordinarily read. Furthermore, it measures rate of reading alone, whereas tests in which the reading material is interspersed with questions measure rate of reading plus rate of answering the questions.

The kind of rate material to use.- Although different kinds of material require different reading rates, it appears that there is marked correlation between rate of reading each type of material and rate of reading other types. For instance, ninety-two pupils in five modern-history classes of the University High School read three selections representing three types of material. The first selection was descriptive, the second was narrative, and the third was expository. The correlations between the rates of reading the different kinds of material were as follows:

	<u>r</u>	<u>P.E.</u>
Descriptive passage with narrative passage	.711 $\pm$	.035
Descriptive passage with expository passage	.691 $\pm$	.037
Narrative passage with expository passage	.729 $\pm$	.033

The preceding correlations are quite high and indicate that there is a decided relationship among rates of reading narrative, descriptive, and expository material.

Since pupils tend to maintain the same relative positions in reading different kinds of material, it may be concluded that the reading matter in a rate test need not be the type most frequently encountered in school, but rather that it should be selected so that it will be as interesting as possible for the individuals to whom it is administered. The main consideration is to so select and arrange it that it will catch the pupil's interest quickly and hold it in order that, if possible, he will forget that he is being tested and will read in the way that he would read the selection for his own pleasure. It was decided, therefore, to begin the experimental work with narratives that contained much action.

The length of the test.- There is much difference in the

length of various standardized tests of reading rate. Some test makers, as Sangren, have used one minute; others, as Gray, have used one hundred seconds; still others, as Stone, have used an entire selection of considerable length and have let the time in which the pupils finish reading be the variable. As there was no experimental evidence concerning the relationship between the length of a rate test and its reliability, two trial forms of a rate test were devised and were used in studying the reliability of rate scores based on different time intervals. Experimentation was carried on with time intervals of one hundred seconds, two hundred seconds, and the length of time required to read an entire selection.

Technique of devising the rate test.- An attempt was made to secure a comprehending rate of reading on the part of all pupils through (1) careful selection of interesting reading matter, (2) arrangement of the test material to emphasize comprehension rather than rate, and (3) controlling the conditions of beginning the test by avoiding the use of "go," "start," or any other term that may be associated with speed tests.

The trial forms contained reading material about Indian games. They were set up after introductory experimentation with a few pupils had indicated that they were about equal in difficulty. Both forms were 1200 words in length.

A new starting technique was tried out in the preliminary forms of the test. The material was so arranged that it began near the bottom of the first page. In administering the test, the examiner read the few lines on the first page aloud while the pupils read them silently, and when the last word was reached the pupils turned to the next page and continued to read silently. Timing was begun as the children turned the page. In this way, the pupils were started on the selection before the real test was begun. Thus, individual differences in quickness of starting were eliminated.

The rest of the space on the first page, except that taken up with the information blanks at the top, was utilized for the questions which were to be answered after the selection had been read. The questions were printed upside down so that the children could not read them before they started the test, but could see that they were there and would know that they would be required to answer them. The purpose of this device was to create in the pupils at the outset a feeling that the test was primarily one of comprehension rather than speed, thus stimulating them to read at a comprehending rate.

The administration of the test.- In giving the test the examiner asked the pupils to mark the place they were reading at

the end of one hundred seconds and at the end of two hundred seconds and also had them record the time used in reading the entire selection. A card system was used to inform the pupils concerning the number of seconds which had elapsed when they finished reading.

Experimental results secured with the trial rate test.-

The two forms of the preliminary rate test were administered to two groups of pupils at the junior-high-school level. One group contained forty-seven pupils and the other contained fifty-eight pupils. The reading rate of the pupils was determined in words per second. Correlations were found between the rates of reading the two forms when the pupils were timed for one hundred seconds, two hundred seconds, and while reading the entire selection. The findings may be summarized as follows:

1. The technique used secured rate scores which were quite reliable when compared with the reliability of standardized reading tests. Most of the reliability coefficients were at least .80. This was significantly higher than the reliability coefficients reported in Chapter II for the rate score on the Monroe Standardized Silent Reading Test and the rate test in the Sangren-Woody Reading Test.¹ It was also higher than the mean of six reliability coefficients reported by Gates² for the rate score on the Courtis Silent Reading Test.

2. Of the various time intervals used, the interval of two hundred seconds seemed to be the best. This length of time resulted in scores that were somewhat more reliable than those secured in a shorter time interval. It also appeared to yield as reliable results as timing the whole selection. It was, therefore, preferable to the latter procedure as it was easier to administer.

3. Twelve hundred words were a satisfactory length for a rate test in which the interval for timing the reading was two hundred seconds. Only two of the 105 pupils who read the preliminary forms of the test finished the entire selection before the signal to mark was given, yet the slowest readers finished within ten minutes, thus keeping the test from being unduly prolonged.

Construction of the four final forms of the test of rate of reading.- As some difficulty was encountered in securing material comparable in difficulty to the trial forms of the rate test, these forms were discarded and the final forms of the test were based on four articles written by Samuel Scoville, Jr., and pub-

¹See pp. 7 and 11.

²A. I. Gates, "An Experimental Study of Reading and Reading Tests," Journal of Educational Psychology, XII (1921), 445, 449.

lished in the Nature Magazine.¹ These articles were especially well suited to the purposes of a rate test for the following reasons:

1. They dealt with wild animal life, a field which is interesting to children of all ages.
2. They were written in story form and contained much action. The thrilling nature of the stories tended to sustain attention from the beginning.
3. The vocabulary, sentence structure, and organization of all four stories presented marked similarities. This fact tended to cause an individual to read all of them at about the same rate.

Permission was obtained from the author to use the material and the following steps were taken to prepare the stories for use in a test of rate of reading.

1. Each story was revised and reduced to a length of exactly 1200 words.
2. Ten multiple-choice questions were made for each form.
3. A set of written directions was prepared.
4. The material was tried out in a preliminary way with a few individuals and minor revisions were made on the basis of their performance. In particular, the story of the Pangolin, which appeared to be read somewhat faster than the other forms, was made slightly harder by changes in its vocabulary.

5. Each of the four forms of the test was arranged in mimeographed form so that the directions were on the first page, the questions were printed upside down on the second page,² the title of the story and the first two or three lines appeared at the bottom of the second page, and the test material was on the third and fourth pages.³

The first experimental use of the four forms with a group of pupils.- The various forms of the rate test were administered to eighty-one pupils in the seventh grade of the University High School in October, 1930. The pupils were tested in three groups, which fact made it possible to alternate the forms and minimize the

¹Samuel Scoville, Jr., "Nanook," Nature Magazine, XIII, No. 2 (February, 1929), 88-90; "Yellow Death," ibid., XIII, No. 1 (January, 1929), 19-21; "Pangolin," ibid., IX, No. 3 (March, 1927), 153-54; "The Diamondback," ibid., XIV, No. 1, (July, 1929), 143-45.

²The reason for this arrangement of the questions is explained on page 21.

³A copy of Form 1 of this test is shown in the Appendix.

effect of practice upon the results.

The intercorrelations among the various forms of the rate test ranged in value from .731 to .881. Four of them were above .80. In general, the correlations were somewhat higher than those which were secured with the two preliminary forms. It has already been stated that the former coefficients were considerably higher than those which had been found for certain other tests. It was therefore concluded tentatively that the new forms of the rate test showed promising evidence of a very satisfactory reliability.

The means and the standard deviations of the rate scores in words per hundred seconds, computed from the amount read during two hundred seconds, are shown in Table IV.

TABLE IV

MEANS AND STANDARD DEVIATIONS OF THE SCORES OF SEVENTY-TWO PUPILS
IN THE UNIVERSITY HIGH SCHOOL ON THE FOUR FORMS
OF THE RATE TEST

Test	Mean Rate	P.E.m	S. D.
Form 1	388	8.16	106.50
Form 2	385	7.37	97.25
Form 3	381	7.16	95.00
Form 4	400	7.76	102.00

The preceding table shows that Form 1, Form 2, and Form 3 of the test of rate of reading were read at approximately the same average rate by the group of University High School pupils and that Form 4 was read slightly faster than the other forms. The standard deviations of the scores on the various forms of the test did not differ greatly.

In general, the scores made by this group of pupils indicated that the four forms of the test of rate of reading could be placed on a comparable basis by a slight revision of Form 4.

Later experimental use of the rate test.¹- Subsequent ex-

¹At two points in the account of the work of making a test of rate of reading mention was made of comprehension questions based on the reading material in the rate test. Discussion of the questions was purposely omitted from this section. The questions will be taken up in the last section of the chapter, which deals with the construction of the comprehension test.

perimentation, in which several hundred pupils in the public schools of Chicago, Oak Park, and Maywood, Illinois, served as subjects, verified the results with the University High School pupils. It showed that the first three forms were comparable and that Form 4 was read more rapidly than the other forms. Changes were then made in the vocabulary of the last form which reduced the average speed of reading the story to that of the first three forms. These data will be presented in Chapter IV.

Construction of the Vocabulary Test

The existing vocabulary tests.- The vocabulary tests that are suitable for use at the junior-high-school level may be classified into two general types, which are these:

1. A short, timed test, containing words that are scaled from very easy to quite difficult. The word-meaning part of the Sangren-Woody Reading Test is an illustration of this type.
2. A longer test, not necessarily scaled, which all pupils are allowed to finish. This type is represented by the Inglis Test of English Vocabulary.

An analysis of these types of tests showed that there was need for a short, reliable vocabulary test which would contain words selected for use at the junior-high-school level and presented to the pupil in context.

Preliminary problems in the making of the vocabulary test.- Before the actual construction of the vocabulary test was undertaken, it was necessary to settle the following questions:

1. From what source should the words be taken?
2. How long should the test be?
3. Should a time limit be placed on the test?

The source of the words.- It was reasoned that a test which was to be used in the junior high school should consist of words found frequently in reading materials and yet should not contain exceptionally easy words. Pitkin¹ had published a list of 1771 words drawn from the less common thousands of The Teacher's Word Book² which contains the 10,000 words which occurred most often in a word count of reading materials. This list appeared to be a pertinent source for a reading vocabulary test since the words contained in it, although not unusually easy, had been found to occur

¹Walter B. Pitkin, The Art of Rapid Silent Reading, pp. 64-73. New York: McGraw-Hill Book Company, 1929.

²E. L. Thorndike, The Teacher's Word Book. New York: Teachers College, Columbia University, 1921.

in written discourse so frequently that pupils would be quite certain to meet them in their reading. As Pitkin's list was sufficiently long to provide a reservoir for a vocabulary test of several forms, it was chosen as the source from which to draw words for the tentative forms of the test.

The length of the test.- Preliminary experimentation carried on mainly by Ivan A. Booker and to a limited extent by the writer showed that fifty-word tests drawn from the Inglis Test of English Vocabulary had a reliability of slightly more than .80. This coefficient of reliability is fairly high. It was decided, therefore, to devise two forms of a vocabulary test, each containing fifty words and to determine their reliability experimentally.

The time limit.- Some vocabulary tests have very brief time limits; in others, the amount of time allowed is so generous that practically all pupils can finish the test. These types of vocabulary tests measure somewhat different things as the correlation between them is quite low.¹

It appears that vocabulary tests that allow pupils to finish are based on the sounder principles as the ultimate purpose of measuring a pupil's vocabulary is to discover whether or not he has a wide knowledge of the meaning of words, rather than to find his rate of working. For this reason, it was decided to adjust the time limit so that most junior-high-school pupils would be able to do the entire test. After some experimentation, it was found that the most desirable time interval for a fifty-word test was fourteen minutes.

Technique of drawing up the test.- The first step in the construction of the vocabulary test was to draw up two preliminary forms each consisting of more than the needed fifty words in order that words which experimentation showed to be unsatisfactory might be eliminated. Accordingly, two forms of a test each containing seventy words taken at regular intervals from Pitkin's list were constructed. Each word was used in a simple sentence or phrase and five words, one of which was similar in meaning to the test word were placed in a line after the sentence and were numbered from one to five. An attempt was made to select synonyms which were more common than the test words. In a few cases it is doubtful if this objective was achieved.

Technique of administration.- The test was found to be very

¹The correlation between the word-meaning part of the Sangren-Woody Reading Test and the Inglis Test of English Vocabulary, based on the scores of eighty-four pupils in the seventh grade of the University High School, was only .57 $\pm$.05.

easy to give as it was possible to make the directions so simple that the pupils could understand them after reading them through once with the examiner. The following directions were printed at the top of the first page:

"DIRECTIONS: In each of the exercises below you are to select that one of the five words or phrases which most nearly corresponds in meaning to the word which is underlined in the sample sentence or expression. Draw a line under the word thus selected and place its number in the parenthesis at the right. Notice the example below:

"EXAMPLE: Let us try to appease him.

(1)allure (2)affect (3)pacify (4)repulse (5)convince . . (3)"

The plan of having the pupils place the number of the word selected at the right of the page greatly facilitated the scoring of the test.

Trying out the preliminary test.- The test was administered to seventy-eight pupils in the University High School during the spring semester, 1930. The pupils were allowed all the time they needed to complete the test. The results indicated (1) that there was no significant difference in the difficulty of the two forms; (2) that the test was neither too hard nor too easy for junior-high-school pupils; (3) that the correlation between the forms was high (.911 for the seventy-eight pupils tested); and (4) that the test showed high correlation (.802) with the long and carefully constructed Inglis Test of English Vocabulary. It was concluded tentatively that the new test appeared to be a reliable and valid test of the vocabulary of pupils at the junior-high-school level.

Reducing the seventy-word tests to fifty-word tests.-

Extensive research studies have shown that questions which have a difficulty value of about 50 per cent, that is, questions on which approximately 50 per cent of the subjects succeed are somewhat more valid than questions at the extreme ends of the scale.¹ Nearly all the very easy and very hard items, therefore, were eliminated from the vocabulary test. With a few exceptions, all the test items retained were answered correctly by more than 20 per cent and less than 80 per cent of the seventh-grade pupils. In reaching a decision about which words in the upper and lower ends of the

¹Thelma Gwinn Thurstone, "The Relation between a Test and its Diagnostic Value." Doctor's Dissertation, University of Chicago Libraries, University of Chicago, 1926;

Clarence Sansom, "The Diagnostic Value of a Test in Relation to Its Difficulty." Doctor's Dissertation, School of Education, University of Chicago, 1929.

scale to eliminate, the desirability of maintaining a balance between the two forms was kept in mind.

Experimentation with the revised test.- The two fifty-word forms of the vocabulary test were administered, in May, 1930, with the regular tests of the entrance battery, to 133 pupils who were to enter the University High School the following autumn. The mean scores made by these pupils on Form 1 and Form 2 were 22.9 and 22.3, respectively. The difference between the means and the probable error of the difference was as follows:

$$M_1 - M_2 = 0.6 \pm 0.26.$$

There was no certainty that the difference between the means was significant as it was less than three times its probable error.

Likewise, there was no marked difference between the distributions of scores on the two forms. The standard deviation of the scores on Form 1 was 10.66 and that of the scores on Form 2 was 10.22.

The reliability of the fifty-word test.- The correlations of the scores on Form 1 and the scores on Form 2 of the vocabulary test are shown in Table V.

TABLE V

RELIABILITY COEFFICIENTS FOR THE FIFTY-WORD VOCABULARY TEST

Groups Used	Number of Cases	r	P.E.
Pupils from public schools	57	.915 $\pm$	.014
Pupils from University Elementary School (Sixth grade)	76	.864 $\pm$	.017
Both Groups	133	.908 $\pm$	.011

Two of the correlations between Form 1 and Form 2 were above .90. The correlation computed with the scores of the fifty-seven public-school pupils was as high as the one secured with the seventy-word forms. It appears that the culling out of very easy and very difficult words from the longer forms of the test did not greatly reduce the reliability of the test.

The validity of the fifty-word test.- The validity of this vocabulary test was investigated by correlating the scores made by the two groups of pupils with their scores on the Inglis Test of English Vocabulary. In general, the correlations confirmed the finding made with the seventy-word preliminary forms to the effect

that there was close agreement between the results of the new test and the results of the Inglis test. The data on which this statement is based will be presented in Chapter IV.

Construction of Form 3 and Form 4 of the vocabulary test.- The next step was the making of two more forms of the vocabulary test that were equivalent in difficulty to Form 1 and Form 2. The procedure employed in the construction of Form 3 and Form 4 was so nearly identical with that used in making the first two forms that a detailed description of it would be valueless.

The preliminary experimentation was carried on with ninth-grade classes in the Englewood High School, Chicago, Illinois. When the last two forms of the test were ready all four forms were tried out with pupils in the Pleasantville, Indiana, High School. No marked differences between the tests were discovered, but a few minor changes were made on the basis of the results.

Experimental use of the revised forms.¹- The four forms of the vocabulary test were then administered to more than two hundred pupils in the ninth grade of the Parker Junior High School, Chicago, Illinois. Means and reliability coefficients found from the scores will be presented in the next chapter, which deals with the evaluation of the entire test.

Construction of the Comprehension Test

The types of comprehension tests included in this battery.- Comprehension in reading consists of an exceedingly complex set of activities. In most of the older tests no attempt was made to measure different kinds of comprehension, but there is a distinct trend in the newer reading tests toward the measurement of more than one aspect of the whole complex process. For instance, the Sangren-Woody Reading Test contains, in addition to sub-tests for vocabulary and rate of reading, the following sub-tests for comprehension: facts, total meaning, central thought, following directions, and organization; the Gates Silent Reading Test measures reading to appreciate the general significance, reading to predict the outcome of given events, reading to understand precise directions, and reading to note details; and the Haggerty Reading Examination, Sigma III, measures sentence meaning and paragraph meaning, in addition to vocabulary.

It is not claimed for any of these newer tests that they represent anything more than a sampling of the multifarious activities which make up comprehension. Even were it possible to construct a comprehension test based upon a complete analysis of

¹See the Appendix for a copy of Form 1.

reading ability, such a test would necessarily be extremely long and unwieldy, if it were reliable enough to be of any real value in diagnostic work. Therefore, one who wishes to make a practicable test in this field can but measure as many of the more important phases of comprehension as will lend themselves to treatment within the time limit that can be given to the test.

There are two phases of the reading comprehension of junior-high-school pupils about which the teacher needs especially to know. In the first place, when a pupil has read a selection, the teacher needs to know whether he has been able to grasp the essence of it. Second, the teacher needs to know how much power of comprehension each pupil has. Both phases of reading are especially important in the classroom activities of the school. It will be shown later in this study that the ability to answer questions based on material read and the ability to grasp the thought of the more difficult reading material are each correlated with school success.

That there is widespread recognition of the importance of ability to read and recall the essential features of what has been read and of ability to understand difficult material is shown by the fact that there are already several standardized tests for each of these aspects of comprehension. All these tests, however, are separate and distinct from each other. In this study, an additional step is taken by incorporating tests for these aspects of comprehension into a single battery of reading tests.

The part of the comprehension test designed to measure the ability of the pupils to answer questions after continuous reading.- The test of rate of reading, which was described in an earlier section of this chapter, became the basis of the first part of the comprehension test. It will be recalled that the rate test was followed by a group of questions printed upside down just preceding the point where the reading material started. When the test was administered, the attention of the pupils was directed to the questions before they commenced to read and they were informed that they were to read the selection and then turn back to the questions and answer them. The questions were thus made to serve a double function: (1) they tended to create in the minds of the pupils a belief that the primary purpose of the test was to measure understanding rather than rate, and thus acted as a check upon erratic individuals who, otherwise, might have skimmed through the test at excessive speed in order to finish ahead of the rest of the group; and (2) they were used to measure the ability of the pupils to understand what they had read.

Preliminary experiment with true-false questions.- The first questions tried out were the true-false type. A large number

of them could be put into a limited space and could be easily scored, but they were found to be of doubtful reliability. As studies have shown that, in general, true-false tests are less reliable than those of the multiple-response type,¹ it was decided to discard the true-false statements and to experiment with five-response questions.

Experimentation with multiple-response questions.- Ten five-response questions were devised for each of the four forms of the rate test. The following examples taken from Form 1 of the test, based on the story "Nanook and her Cub," illustrate the manner in which the questions were presented to the student.

- "1. The cub whimpered because (1) he was cold
(2) he was tired (3) he was hungry (4) he was lost
(5) he was afraid ()

- "2. The first kill the bear made was
(1) a pair of ruffed grouse (2) a pair of ptarmigan
(3) a white rabbit (4) a young deer (5) a musk ox . . . ()"

The pupils were instructed to underline the word or phrase which should be used to complete the statement and to place its number in the parenthesis at the end of the line.

The intercorrelations of the scores made on the four forms of the test by a group of eighty-one seventh-grade pupils in the University High School ranged from .592 to .774. The mean of the six reliability coefficients was .654. This was .227 higher than the reliability coefficient for the scores of a comparable group of pupils on questions of the true-false type.

Although the correlations between the various forms of this sub-test were not very high, they were probably as high as could be expected, in view of the fact that this part of the test contained only ten questions. The correlations were approximately the same size as the reliability coefficients presented in Chapter II for the Monroe Standardized Silent Reading Test and the comprehension parts of the Sangren-Woody Reading Test.²

Means and standard deviations of scores.- The means and standard deviations of the scores of the preceding group of seventh-grade pupils indicated that Forms 1, 2, and 4 were nearly equivalent in difficulty, although Form 1 appeared to be slightly harder than the other forms. Form 3, however, was much easier than the other three forms and the standard deviation of the scores was smaller. It obviously required decided revision.

Revision of the comprehension questions and experimental use of the revised test.- One of the harder questions in Form 1 was

¹Ruch, op. cit., pp. 296, 301, 302.

²See Tables I and III for comparative data.

changed slightly in an effort to lessen its difficulty and to make this form equal in difficulty to Forms 2 and 4. The questions in Form 3 were subjected to a critical analysis and were thoroughly revised. Informal trial with a few individuals indicated that this form then approached the other forms in difficulty.

Forms 1 and 2 of the revised comprehension questions were administered to 302 public-school pupils in the ninth grade and Forms 3 and 4 were given to 327 pupils in the same grade. Form 1 and Form 4 were also given to fifty-nine ninth-grade pupils. The results indicated that the forms were comparable in difficulty and that the questions were well adapted to the ability level of the pupils who served as subjects. The data supporting this point will be presented in Chapter IV.

Part II of the comprehension test designed to measure power of reading.- A power test presupposes that the influence of the time factor is reduced to a minimum; that is, it should be so constructed as to length that all, or nearly all, pupils can complete the test within the time limit. Therefore, before constructing such a test, one needs to know the amount of time that is to be given to it.

In the battery of tests constructed in this study, about thirty minutes were required to administer the vocabulary test and the rate test, together with the comprehension questions following it. Since it seemed desirable to have the entire battery of tests come within a fifty-minute period, approximately twenty minutes were left for the second part of the comprehension test.

Technique of making a trial form of the power test of comprehension.- Eight paragraphs were chosen from a school reader, a book of model paragraphs, and a textbook of philosophy, and were revised and adapted for purposes of a comprehension test. Three questions were formulated for each of seven of the paragraphs and four questions were devised for the other paragraph, making twenty-five questions in all.

The paragraphs and their questions were placed in order of difficulty by five judges, all of whom showed close agreement in the relative rank assigned. The composite rank assigned by the judges was used in determining the tentative order in which the paragraphs were placed in the test. The test was set up in mimeographed form, beginning with the paragraph which the judges thought easiest and ending with the one which they considered most difficult. A set of directions and a sample exercise preceded the test.

Experimental results secured with the trial form of the comprehension test.- The test was administered to fifty-eight children who were prepared to enter the University High School from outside schools and to seventy-five pupils in the sixth grade of

the University Laboratory Schools. The test was not timed and the subjects were permitted to reread the paragraphs as often as was necessary before answering the questions.

The results of the test were correlated with scores made by the same pupils on the Thorndike-McCall Reading Scale and the Monroe Standardized Silent Reading Test. The correlation coefficients were not very high (approximately .60). Nevertheless, the new test correlated somewhat higher with each of the standardized tests than the standardized tests correlated with each other.

The mean scores made on the test by the sixth-grade pupils and by the entering high-school students were 19.2 and 20.0 respectively. It is evident that the pupils found the test relatively easy. The average percentage of correct answers for the whole test was 80 per cent.

Revision of the test and construction of parallel forms.-

The next steps in the development of the power test were to increase the difficulty of the trial test and to make duplicate forms of the test. The following procedure was used to achieve those objectives:

1. An inspection of the results of the preliminary test showed that the odd-numbered paragraphs were not greatly different in difficulty from the even-numbered ones, except that Paragraph 3 was considerably easier than Paragraph 4. Accordingly, the third paragraph was discarded and one which was found to be harder was substituted in its place.

2. The preliminary test was then divided into halves, so that Paragraphs 1, 3, 5, and 7 formed one half and Paragraphs 2, 4, 6, and 8 comprised the other half.

3. Three additional paragraphs were added to each of the preceding parts. Each of the sets of new paragraphs consisted of a paragraph of narrative material adapted from a school reader, a paragraph of semi-political character from a daily newspaper, and a paragraph adapted from a book of philosophy. The last two types of paragraphs were relatively difficult.

4. Twenty-five questions were made out for each set of seven paragraphs. Most of the questions on the paragraphs which had been used in the original test were retained. In form, the questions were of two types--brief-answer questions and five-response questions. The majority of the questions dealt with facts, but about 25 per cent of those in each form were designed to test grasp of total meaning of a paragraph.

5. The two sets of seven paragraphs and twenty-five questions were called Form 1 and Form 2 of a Test of Comprehension in Silent Reading. These two forms served as a model in the construction of Form 3 and Form 4 of the test. The latter forms, likewise,

contained seven paragraphs and twenty-five questions, arranged roughly in order of difficulty.

6. These preliminary forms were submitted to three judges for criticism and were revised on the basis of their criticism.

The first trial and revision of the four forms.- All forms of the test were administered to a class of twenty-five pupils in the ninth grade of the University High School. Although the various forms appeared to be about equally difficult, two adverse facts were apparent in the results: (1) several of the questions were so easy that the responses to them were right in practically all the cases; and (2) the test, even though it was shorter by one paragraph than the preliminary test, was still so long that the slowest readers could not cover it in twenty minutes.

Accordingly, each form was revised and reduced to six paragraphs and twenty questions. Two criteria were used in revising the tests: (1) the difficulty of the various questions, and (2) the validity of the questions. The validity of the questions was studied by noting the pupils who succeeded with a question did better with the test as a whole than the pupils who failed on a question. When there was considerable doubt about a test item, the biserial r of the item with the whole test was computed.¹

Trying out Form 1 and Form 2 after revision.- The experimental work with the power test thus far had been with the University High School pupils. In order to find out whether or not the test was well adapted for use with public-school pupils, Form 1 and Form 2 of the test were administered to eighty-one pupils in Grade IX of the Englewood High School, Chicago, Illinois. The results of the testing pointed to the following conclusions:

1. The forms of the test could be made equivalent in difficulty with slight revisions.

2. The distribution of scores on Form 1 was much like the distribution of scores on Form 2. The standard deviations were, respectively, 3.14 and 3.17.

3. The test seemed to be well adapted in difficulty for public-school pupils at the junior-high-school level. About half

¹The formula for biserial r is

$$r_{bis} = (\bar{Y}_2 - \bar{Y}_1 / \sigma_Y) (pq/z),$$

where $\bar{Y}_1$ and $\bar{Y}_2$ are the means of the two columns, $q = n_1/N$ = proportional part of the cases in the first category, $p = n_2/N$ = proportion of cases in the second category, and z = the ordinate of the normal curve for area p (Holzinger and Mitchell, An Exercise Manual in Statistics, pp. 120-21. Boston: Ginn and Company, 1929).

the answers to the questions were correct. No pupil made a perfect score and no one scored zero on either form of the test.

4. The test was fairly reliable in view of the fact that it contained only twenty questions. The correlation between the forms was

$$r = .737 \pm .034.$$

Further experimentation with the four forms of the power test.- The next step in the development of the power test of comprehension was to administer all four forms to a larger group of junior-high-school pupils and to determine mean scores, standard deviations, reliability coefficients, and validity of individual items. The test was given to several classes in the eighth grade of the Amundsen Junior High School. A number of changes were made in the items of the test on the basis of the results, but those changes were, with one exception, minor ones. The exception was that the questions on Paragraph 5 in Form 4 were found to be so greatly lacking in objectivity that the whole paragraph was discarded and a new one was substituted for it.

Securing a total comprehension score.- The final step in the construction of the comprehension test was to combine the comprehension questions on the rate test with those in the power test for a total comprehension score in each form. The whole comprehension test was tried out with various groups of public-school pupils, and the reliability and the equivalence of the forms were studied.

The Complete Battery of Tests

Assembling the parts of the test into final form.- When it appeared that there were four equivalent forms of each part of the reading test, and that each part met acceptable standards, the various sub-tests were combined into parallel forms of a battery of tests. Each form was assembled in such a manner that the rate test and its comprehension questions were placed first, the vocabulary test second, and the power test of comprehension last.

Experimental use of the complete battery.- In order to obtain data on the reliability and validity of the entire battery of tests, Form 1 and Form 2 were administered to 150 ninth-grade pupils, Form 2 and Form 3 were given to 54 pupils, and Form 1 and Form 4 were administered to 53 ninth-grade pupils. Forty-six minutes of working time were allowed in the administration of the test. The time was distributed as follows: rate test and comprehension questions, twelve minutes; vocabulary test, fourteen minutes; comprehension test, twenty minutes. All pupils except the very slowest workers were able to cover the material in each part within the time limit. The findings will be presented in Chapter IV, which deals with the evaluation of the reading tests.

CHAPTER IV AN EVALUATION OF THE TESTS

The Problem of the Chapter

The purpose of this chapter is to make a critical study of the tests whose construction has been described in an attempt to discover the degree to which they meet the criteria of good tests that were listed in Chapter III.¹ The criteria will be applied to the tests in the order in which they were listed. Thus, those which are more obvious and more readily disposed of will be taken up first, and those which are less obvious and which require more extended statistical treatment will be reserved for the latter part of the chapter.

Ease of Administration²

Factors which facilitate the administration of the test.-

The test is easy to administer for several reasons. First, the battery is so constructed that nearly all pupils can finish it within a class period of fifty minutes. This makes it possible to give the whole test at one sitting. Second, at every point where there is any likelihood that the pupil will go astray, there are printed directions telling him exactly what to do. Third, the test of reading vocabulary and the test of understanding in silent reading are practically self-administering. In fact, in giving those tests to individual pupils, the plan has sometimes been followed of placing the test blank before the pupil without verbal instructions other than a statement that he was to read the directions and do the test without waiting for a starting signal. Invariably, the subject has been able to go ahead without further help. Thus, it appears that after the first part of the test has been completed, the chief function of the examiner is merely that of time-keeper. The pupils are able to work with very little supervision.

Danger points in the administration of the test.- There are a few opportunities for children to go astray while taking the test. The first place is where the pupil turns to page 2 and observes that most of the material on that page is printed upside

¹See page 19.

²This section will be more meaningful if it is read after an inspection of the test shown in the Appendix.

down. There are, however, three safeguards against confusion at that point—the detailed explanation of page 1, an explanatory paragraph at the top of page 2, and verbal directions and explanation which the examiner gives as soon as the pupils have turned to the second page. These are so effective that the arrangement of the material on the second page rarely causes trouble for any of the pupils.

The second point of possible difficulty is at the end of the story where these instructions are found: "Now turn back to page 2 and answer the questions." That sentence is usually sufficient to get pupils to turn at once to the questions. If a child is in doubt, he usually glances about and observes what other pupils are doing and then follows their lead. It may be necessary, occasionally, for the examiner to go to a pupil and show him what to do next, but during the administration of several hundred tests it has not been necessary to do that in more than half a dozen cases.

Ease and Speed of Scoring

How the scoring is done.— The scoring of all parts of the test except the power test of comprehension is entirely mechanical and can be done easily and quickly. The scoring of the power test requires more time than any of the other parts of the reading test. The reason for this is that a large percentage of the questions are the brief-answer type, and, consequently, the devices used to simplify the scoring of the other parts of the test are not feasible in this part.

Scoring time for the entire test.— Four graders, one of whom was familiar with the test, were timed while they scored 655 Form-one test papers. The amount of time required was, in round numbers, 2700 minutes, or approximately four minutes for each paper. At that rate, a teacher will spend about two hours in scoring the papers of a class of average size. The amount of time required is not excessive, in comparison with other standardized tests of similar length.

Objectivity of the Tests

The battery of tests constructed in this study stands fairly high in objectivity. The rate test is entirely objective. If a pupil reads 600 words in 200 seconds, all scorers must agree that he reads at a rate of three words a second. Likewise, the five-response items used in testing comprehension of continuous material and in the vocabulary test are thoroughly objective. A teacher might be disposed to question the value of some of the test items, but she must admit that each one permits but one correct

response. Consequently, all teachers will be obliged to score all items in the same way.

Although the power test contains a number of questions of the brief-answer type, most of the questions have only one correct answer. This test is almost, if not quite, as objective as the other parts of the test.

Equivalence of the Forms of the Tests

The equivalence of the forms of the rate test.- The relative size of the means and standard deviations of the scores made on the forms of the rate test by three groups of pupils is shown in Table VI. Group I contained 309 ninth-grade pupils; Group II contained 53 pupils; and Group III comprised 109 ninth-grade pupils.

TABLE VI

MEANS, DIFFERENCES BETWEEN MEANS, PROBABLE ERRORS OF DIFFERENCES BETWEEN MEANS, AND STANDARD DEVIATIONS OF SCORES MADE BY THREE GROUPS OF PUPILS ON VARIOUS FORMS OF THE RATE TEST

Form	Group	Mean Score	Difference between Means	P.E. of Difference*	Difference = ? x P.E.	S.D. of Scores
1	I	34.30				7.55
2	I	34.00	0.30	0.13	2.3	7.05
2	II	34.20				10.75
3	II	36.00	1.80	0.52	3.5	11.60
1	III	37.30				8.97
4	III	37.20	0.10	0.36	0.3	8.65

*The formula for finding the probable error of the difference between correlated means is:

$$P.E.m_1 - m_2 = \sqrt{(P.E.m_1)^2 + (P.E.m_2)^2 - 2r_{12}P.E.m_1P.E.m_2}$$

(Karl J. Holzinger, Statistical Methods for Students in Education, Formula 100.)

The preceding table is read as follows: The mean of the scores made by Group I on Form 1 of the rate test is 34.30; the mean of the scores made by the same group on Form 2 of the test is 34.00; the difference between the means is 0.30; the probable error of the difference is 0.13; and the difference is 2.3 times the

probable error. The standard deviations of the scores on Form 1 and Form 2 are, respectively, 7.55 and 7.05.

The table shows that the forms of the test of rate of reading were read at approximately the same average rate. Form 3 was read slightly faster than Form 2, but the differences between the other forms do not appear to be significant. The general rule is that a statistical constant must be at least three or four times its probable error to be significant.¹

The equivalence of the forms of the vocabulary test.- The means and standard deviations of the scores made on the vocabulary test by 208 pupils in the ninth grade of the Parker Junior High School, Chicago, Illinois, in December, 1930, are shown in Table VII.

TABLE VII

MEANS, DIFFERENCES BETWEEN MEANS, PROBABLE ERRORS OF DIFFERENCES BETWEEN MEANS, AND STANDARD DEVIATIONS OF SCORES MADE BY 208 NINTH-GRADE PUPILS ON THE FOUR FORMS OF THE VOCABULARY TEST

	Mean Score	Difference between Means	P.E. of Differ- ence	Difference = ? x P.E.	S.D. of Scores
Form 1	24.40	--	--	--	9.00
Form 2	24.40	--	--	--	9.25
Form 3	24.57	--	--	--	8.75
Form 4	24.05	--	--	--	8.15
M1 - M2	--	0	0	0	--
M3 - M1	--	0.17	0.21	0.8	--
M1 - M4	--	0.35	0.24	1.5	--
M3 - M2	--	0.17	0.23	0.7	--
M2 - M4	--	0.35	0.25	1.4	--
M3 - M4	--	0.52	0.23	2.3	--

The foregoing table shows that the forms of the vocabulary test are practically equal in difficulty. The differences among the means of the scores on Form 1, Form 2, and Form 3 are very small. In fact, the means of the scores made by this group of pupils on Form 1 and Form 2 are identical. Form 4 appears to be

¹Karl J. Holzinger and Blythe C. Mitchell, Exercise Manual in Statistics, p. 90. Boston: Ginn and Company, 1929.

slightly harder than the other forms, but the differences between the mean score made on it and those made on the first three forms are quite probably due to chance factors.

The differences in the standard deviations of the scores on Form 1, Form 2, and Form 3 are small. The scores on Form 4 show considerably less variability than those on the other forms.

In general, it may be concluded that the scores on the four forms of the vocabulary test are directly comparable, although it is probable that the chances of making a very high or a very low score on Form 4 are not quite as great as on the other forms of the test.

Mean scores on Part I and Part II of the Comprehension Test.- The mean scores made by various groups of pupils on the different forms of each part of the comprehension test are shown in Table VIII. The groups used were as follows: Part I, Forms 1 and 2, 302 ninth-grade pupils; Forms 3 and 4, 327 ninth-grade pupils; Forms 1 and 4, 59 ninth-grade pupils; Part II, all forms, 103 eighth-grade pupils.

TABLE VIII

MEANS, DIFFERENCES BETWEEN MEANS, AND PROBABLE ERRORS OF DIFFERENCES BETWEEN MEANS OF SCORES MADE BY PUPILS AT THE JUNIOR-HIGH-SCHOOL LEVEL ON THE VARIOUS FORMS OF PART I AND PART II OF THE COMPREHENSION TEST

	Mean Score	Differ- ence between Means	P.E. of Differ- ence	Differ- ence $\pm$? x P.E.
Comprehension, Part I				
Form 1	5.31			
Form 2	5.37	0.06	0.07	1-
Form 3	6.80			
Form 4	6.92	0.12	0.06	2
Form 1	5.25			
Form 4	4.90	0.35	0.19	2-
Comprehension, Part II				
Form 1	11.54	--	--	--
Form 2	11.04	--	--	--
Form 3	10.88	--	--	--
Form 4	10.76	--	--	--
M ¹ - M ²	--	0.50	0.15	3.3

TABLE VIII (Continued)

	Mean Score	Differ- ence between Means	P.E. of Differ- ence	Differ- ence = ? x P.E.
M ¹ - M ³	--	0.66	0.15	4.4
M ¹ - M ⁴	--	0.78	0.15	5.2
M ² - M ³	--	0.16	0.15	1.1
M ² - M ⁴	--	0.28	0.15	1.8
M ³ - M ⁴	--	0.12	0.14	0.9

None of the differences in the means of the scores on the forms of Part I administered to the same group are significant.

The data for Part II indicate that the differences between the mean scores made on Form 1 and the mean scores made on the other forms are significant. The differences are from three to five times the probable errors of the differences. The differences between the mean scores on Form 2, Form 3, and Form 4 are not significant.

When it was discovered that Form 1 was somewhat easier than the other forms of the power test, an effort was made to bring the forms closer together through small changes in some of the test items. That the objective was accomplished is indicated by the following mean scores which were secured, after slight revisions had been made, by administering Form 1 and Form 4 to fifty-three ninth-grade pupils.

Form 1 12.30
Form 4 12.64

The difference between the means, .34, is insignificant, as it is only 1.3 times the probable error of the difference.

The equivalence of the total comprehension scores.- It was pointed out earlier that when the various tests are arranged in a battery, the scores on the two types of comprehension tests are combined for a total comprehension score and that the comprehension score is then multiplied by two so that comprehension will contribute approximately as much to the composite score on the whole battery as do rate and vocabulary. Data concerning the equivalence of the total comprehension scores on the various forms of the test were gathered by administering pairs of forms to three groups of pupils. Group I comprised 150 ninth-grade pupils in the Medill Junior High School and the Proviso Township High School; Group II contained 54 pupils in the Medill Junior High School and the

University High School; Group III contained 53 ninth-grade pupils in the Oak Park and River Forest Township High School. The means and standard deviations of the scores made on the tests are shown in Table IX.

TABLE IX

MEANS, DIFFERENCES BETWEEN MEANS, PROBABLE ERRORS OF DIFFERENCES BETWEEN MEANS, AND STANDARD DEVIATIONS OF SCORES MADE BY THREE GROUPS OF PUPILS ON VARIOUS FORMS OF THE COMBINED COMPREHENSION SCORE

Form	Group	Mean Score	Difference between Means	P.E. of Difference	Difference = ? x P.E.	S.D. of Scores
1	I	29.5				10.5
2	I	31.2	1.7	0.35	4.9	10.1
2	II	35.1				12.9
3	II	33.9	1.2	0.66	1.8	13.9
1	III	38.5				9.0
4	III	37.9	0.6	0.54	1.1	7.5

The differences between the means shown in Table IX are small. Form 2 seems to be slightly easier than the other forms of the test. The standard deviations of the scores on Form 1 and Form 2 and of the scores on Form 2 and Form 3 are quite similar. The difference between the standard deviation of the scores on Form 1 and that of the scores on Form 4 is one of considerable size, but it is of doubtful significance because of the small number of cases involved.

The equivalence of the composite scores on the whole battery.- As the differences between the scores on the four forms of the various parts of the test battery were found to be small, in most instances, it is logical to assume that the total scores on the forms of the battery are about equivalent. The correctness of such an assumption is verified by the means and standard deviations of the composite scores made by the three groups of pupils that were used in comparing the forms of the comprehension test.

No large differences between the means appear in Table X. None of the differences are as much as four times the probable error of the difference.

TABLE X

MEANS, DIFFERENCES BETWEEN MEANS, PROBABLE ERRORS OF DIFFERENCES BETWEEN MEANS, AND STANDARD DEVIATIONS OF THE SCORES MADE BY THREE GROUPS OF PUPILS ON VARIOUS FORMS OF THE ENTIRE TEST BATTERY

Form	Group	Mean Score	Difference between Means	P.E. of Difference	Difference = ? x P.E.	S.D. of Scores
1	I	80.81				21.93
2	I	82.53	1.72	0.52	3.3	22.28
2	II	94.00				31.00
3	II	93.91	0.09	0.95	0.1	33.27
1	III	106.83				21.47
4	III	104.89	1.94	0.82	2.4	18.57

The distribution of the scores on Form 1 is much like that of the scores on Form 2, and the distribution of the scores on Form 2 is similar to that of the scores on Form 3. The scores on Form 4, however, show considerably less variability than the scores on Form 1. This is consistent with the findings which have already been presented for the various parts of the test battery.¹

In general, the data presented in this section of the study indicate that the forms of the tests are approximately equivalent. While differences exist in the means and in the variability of the scores on the different forms of the tests, all of the differences are small and most of them can be accounted for by sampling fluctuations. It is probable that more extensive use of the test with a larger number of cases will reveal changes which should be made in order to bring the forms closer together, but the facts which have been presented support the conclusion that the forms of the test are now about equal.

¹The wide differences between the standard deviation of the scores made by the second group and those made by the first and third groups probably results from the fact that the second group was composed partly of pupils in a public school and partly of pupils in the University High School. Most of the pupils in the latter school were superior readers, and their scores tended to exceed those of the public-school pupils by a considerable amount.

Tentative Norms

Scores of pupils in one junior high school.- In order to establish tentative norms of achievement for pupils in the seventh, eighth, and ninth grades, Form 1 of the battery of tests was administered to 583 pupils in the Lowther Junior High School, Emporia, Kansas, in September, 1931. Emporia is a typical small mid-western city, with a population of about 15,000, and it is probable that the pupils in the junior high school there are fairly representative of the general population of junior high schools in the country at large.

The means of the scores made on the test by the Emporia pupils are shown in Table XI.

TABLE XI

MEAN SCORES MADE BY THE PUPILS IN THE LOWTHER JUNIOR HIGH SCHOOL,
EMPORIA, KANSAS, ON FORM 1 OF THE SILENT READING TEST
WHICH WAS CONSTRUCTED IN THIS STUDY

Grade	Number of Pupils	Rate	Vocab- ulary	Compre- hension 1	Compre- hension 2	Compre- hension Total	Total Score
7th - Mean . .	198	34.6	12.9	5.2	6.8	24.0	71.6
P.E. . .		0.49	0.27	0.10	0.14	0.43	0.93
8th - Mean . .	190	35.3	16.9	5.6	8.0	27.2	79.5
P.E. . .		0.54	0.37	0.11	0.15	0.46	1.10
9th - Mean . .	195	36.2	19.6	6.3	10.4	33.4	89.2
P.E. . .		0.45	0.42	0.11	0.17	0.49	1.06

Significance of the difference in grade achievement.- The data in Table XI indicate that the tests make fairly satisfactory discrimination between average achievement of pupils in successive junior-high-school grades. The differences in the case of rate of reading are quite small, but this fact may depend upon slow maturity of that function after the seventh grade. The differences in average grade achievement in vocabulary, comprehension, and composite score are apparently significant.

Validity of the Tests

How validity is studied.- In the evaluation of any test possibly the most important single question that can be raised

concerning it relates to whether it measures what it is designed to measure. The validity of the tests constructed in this study was investigated by the following procedures:

1. Each of the tests in the battery was validated by correlating it with the best available criterion or criteria of the particular type of reading ability measured by the test.

2. The degree to which the test as a whole correlates with a somewhat similar, but longer, battery of standardized tests was investigated.

3. The value of the tests in the prognosis of school success was studied.

4. The validity of each of the individual items in the test was computed.

5. The reliability of the forms of the test was found. This indicated validity only in a negative way. A test which is reliable is not necessarily valid, but no test can be valid unless it is also reliable.

The validity of the rate test.- The following steps were used in studying experimentally the validity of the test of rate of reading: (1) several passages which represented different kinds of reading material were selected; (2) a class of twenty-two pupils read the material under conditions in which comprehension was emphasized; (3) the pupils were timed during the reading; and (4) the first three forms of the rate test were administered to the group and the average scores made by the various pupils on the three forms were correlated with their average rate scores on the different kinds of material. The purpose in adopting this procedure for studying the validity of the rate test was to find out whether or not pupils tend to occupy the same relative position within a group in rate of reading the test material that they hold in reading other kinds of material.

The mean scores of each pupil on the three forms of the rate test rather than scores on single forms were used in the correlation because the question under consideration was concerned with the validity of the technique employed in the rate test, and, therefore, it was desirable that fluctuations in performance on individual forms of the test, due to various extraneous factors, should be minimized as much as possible. It was found that

$$r = .812 \pm .05.$$

This correlation is quite high. It furnishes evidence to support a conclusion that the test devised in this study is valid for the measurement of rate of reading.

The validity of the vocabulary test.- The Inglis Test of English Vocabulary is one of the longest and most carefully constructed tests available for use with junior-high-school pupils.

For this reason, it was considered the best criterion for word knowledge that was available. The scores that were made by four groups of pupils on the test constructed in this study were correlated with their scores on the Inglis test. The groups were as follows: (1) fifty-seven public-school pupils; (2) seventy-six pupils in the sixth grade of the University Elementary School; (3) forty-seven public-school pupils; and (4) sixty-one pupils in the sixth grade of the University Elementary School.¹ The first two groups took the University High School entrance tests in May, 1930, and the last two groups took the entrance battery in May, 1931. The tests given in 1930 included Form 1 and Form 2 of the new test, but Form 1, only, was used in 1931. Correlations between scores on the Inglis test and scores on the new vocabulary test are shown in Table XII.

TABLE XII

CORRELATIONS BETWEEN THE CONSTRUCTED VOCABULARY TEST
AND THE INGLIS TEST OF ENGLISH VOCABULARY

Group	Number of Cases	Inglis and Form 1	Inglis and Form 2	Inglis and Both Forms
I	57	.852 $\pm$.024	.866 $\pm$.024	.876 $\pm$.020
II	76	.756 $\pm$.033	.803 $\pm$.028	.800 $\pm$.028
III	47	.856 $\pm$.026	---	---
IV	61	.810 $\pm$.033	---	---

The lowest correlation shown in Table XII is .756 and the highest is .876. The mean of the eight correlation coefficients is .827. It is apparent that there is close agreement between these two tests. If the Inglis test measures knowledge of word meaning accurately, the vocabulary test which was made in this study is, undoubtedly, a valid one.

The new vocabulary test and the Inglis test were also correlated with the word-meaning part of the Stanford Achievement Test. The scores of the pupils in Group IV were used in the correlations. It was found that the correlation of the new vocabulary

¹The Inglis Test of English Vocabulary cannot, ordinarily, be used below the ninth-grade level, unless the pupils are a somewhat selected group, as was true in this case.

test and the Stanford Achievement Word-Meaning Test was $.780 \pm .033$; and that the correlation of the Inglis Vocabulary Test and the Stanford Word-Meaning Test was $.767 \pm .035$.

The vocabulary test made in this study correlates somewhat higher with each of the standardized tests than they correlate with each other, although the differences are small. This is additional evidence in favor of the validity of the new test.

The validity of the comprehension test.- It is difficult to study the validity of a comprehension test because comprehension is not a unitary process. The development of a satisfactory criterion with which to correlate a test of comprehension would involve a complete analysis of the complex whole into its elements, the discovery or creation of a test for each of the elements, and the use of those tests in a weighted composite score or criterion score. It is obvious that such an undertaking would itself constitute an extensive study, and that the final product would depend, in considerable degree, upon subjective judgment.

In the absence of a suitable criterion of comprehension, the correlations among various comprehension tests has been used in estimating the validity of the tests, although they are not conclusive evidence of it. Two comprehension tests might show low correlation with each other, not because of low validity, but because they tested different aspects of comprehension. Nevertheless, if a comprehension test is correlated with several other comprehension tests and low coefficients of correlation are secured with consistency, the validity of the test is open to question. On the other hand, if the correlations tend to be high, the evidence favors the assumption that the test has considerable validity.

The comprehension scores of fifty-four pupils in the sixth grade of the University Laboratory Schools were secured with the Thorndike-McCall Reading Scale, the Monroe Standardized Silent Reading Test, the Paragraph-Meaning part of the Stanford Achievement Test, and the test devised in this study. The scores made on each test were correlated with the scores on the other tests. The correlation coefficients are shown in Table XIII.

The entries in Table XIII indicate that the new comprehension test has a decided tendency to agree closer with the various standardized tests than they agree with each other. The differences between the coefficients are small but consistent. The data scarcely warrant a generalization, but it seems probable that the newly constructed test is at least as valid a test of comprehension as the three widely used standardized tests with which it was compared.

The scores on the three standardized tests were combined

TABLE XIII

CORRELATIONS AMONG VARIOUS COMPREHENSION TESTS

Tests Correlated	<u>r</u>	P.E.
Stanford Achievement Paragraph-Meaning and Monroe Comprehension	.663	± .052
Stanford Achievement Paragraph-Meaning and Thorndike-McCall	.655	± .053
Monroe Comprehension and Thorndike-McCall	.588	± .063
New comprehension test and Stanford Achievement Paragraph-Meaning	.729	± .043
New comprehension test and Monroe Comprehension	.670	± .051
New comprehension test and Thorndike-McCall	.733	± .043

into a crude criterion score by adding the comprehension scores made by each pupil, without weighting the tests. The correlation of the scores on the constructed comprehension test with the criterion scores was $.799 \pm .033$. This is a fairly high coefficient. Although the criterion scores are highly fallible, they are probably somewhat more nearly indicative of the real reading ability of the pupils than are the scores on a single test. The fact that the new comprehension test shows rather high correlation with the criterion suggests that the test is fairly valid for testing comprehension in silent reading.

The validity of the composite score.- It was stated in Chapter II that the selection of pupils in the Sub-Freshman and Freshman classes of the University High School for special guidance in reading has been made on the basis of a battery of standardized tests administered when the pupils were admitted to the school. Four tests which have been used regularly in that battery are the Monroe Standardized Silent Reading Test, the Thorndike-McCall Reading Scale, the Inglis Test of English Vocabulary, and the Otis Self-Administering Test of Mental Ability. Although these tests have consumed more time than was desirable, they have been found to be very serviceable in the identification of inferior readers.

Data concerning the agreement between that series of tests and the new test battery were secured by administering both batteries to fifty-eight pupils who had nearly finished the sixth grade of the University Laboratory Schools. The total scores of the pupils on the battery of entrance tests were correlated with their composite scores on the new tests. The rate score on the

Monroe Standardized Silent Reading Test was not used as it was thought that the use of that score in addition to the comprehension score would give the Monroe test too much weight in the total score. The correlation between the batteries of tests was

$$r = .882 \pm .020.$$

Obviously the scores on the two series of tests were in close agreement. The preceding correlation coefficient is only a little lower than the correlation between two forms of the new test.¹

School marks used as a criterion of reading ability.-

Without question, the success of pupils in class work depends to some extent upon their ability to read. Under a system of education in which textbooks and supplementary reading materials, rather than the teacher's presentation of the subject matter, constitute the chief bases of instruction, children who have not learned to get thought from the printed page readily are at a distinct disadvantage. Possibly assignments at the junior-high-school level are not extensive enough to render rate of reading an important factor in school success, but certainly a wide acquaintance with word-meaning and ability to comprehend written materials are prime essentials in the acquisition of knowledge and understandings. It seems logical, therefore, to assume that if a reading test is a valid one, it will show significant correlation with school success.

One obstacle to the use of the success of pupils in school as a criterion of the validity of a reading test is that there is no infallible index of school success. The marks given by teachers are based so largely upon subjective judgment that they are notoriously untrustworthy. However, they probably constitute the best available data on the degree of success which pupils have achieved in the classroom. If several marks are averaged, doubtless they indicate, in a general way, whether a pupil has been a superior, average, or inferior student. It seems, then, that they possess some value for use in studying the validity of a test of reading.

A large number of studies of the correlation between school marks and the scores on reading tests have been made. Most of the correlation coefficients which have been found are rather low but significant. They are seldom higher than .5. For example, the writer made a study of the correlation of the scores on entrance tests and the school marks of seventy-one seventh-grade pupils, ninety-eight ninth-grade pupils, and ninety-four tenth-grade pupils

¹See the next section of this chapter for comparative data, p. 55.

in the University High School.¹ Scores on the following reading tests, in addition to several intelligence and achievement tests, were used: the Gray Silent Reading Test, the Monroe Standardized Silent Reading Test (rate and comprehension), the Thorndike-McCall Reading Scale, the Burgess Silent Reading Test, and the Inglis Test of English Vocabulary. A total of twenty-two correlation coefficients were computed for scores on various reading tests and school marks. The coefficients ranged in value from .005 to .506. The mean of the twenty-two correlations was .339. The Thorndike-McCall Reading Scale was the only reading test that showed correlations with teachers' marks which were consistently above .40.

In the present study, the scores made by eighty ninth-grade pupils in the Proviso Township High School, Maywood, Illinois, on Form 1 of the new test battery were correlated with the average marks which the pupils received at the end of the year in which the tests were given. Scores made by the same pupils on the Monroe Standardized Silent Reading Test and I.Q.'s secured from their scores on the Otis Self-Administering Test of Mental Ability, which tests were administered as a part of the regular measurement program of the school, were also correlated with average school marks. The marking system was a five-point number scale, of which I was the highest, II was next to highest, et cetera.

The correlations between the school marks and the scores on the various tests are shown in Table XIV.

Neither the rate test constructed in this study nor the rate scores on the Monroe Standardized Silent Reading Test show significant correlation with school marks. A study of the correlation table for school marks and the rate test devised by the writer shows that exceedingly rapid readers made high marks and that very slow readers made low school marks. But, with the exception of the extreme cases just mentioned, the rate scores were not correlated with school success. However, there is little likelihood that this fact furnishes evidence against the validity of either of the rate tests. Rather, it is probable that these students had not yet reached that level in school progress where the amount of extensive reading required of them placed the slow readers at a marked disadvantage. Or it may be that determination and constant application on the part of the slow readers make up for their lack of speed in covering reading material.

¹Unpublished Annual Report of the Remedial Workers in Reading in the University High School, The University of Chicago, 1929-30. Filed in the office of the Research Secretary, The Laboratory Schools, The University of Chicago, Chicago, Illinois.

TABLE XIV

CORRELATIONS OF TEST SCORES WITH AVERAGE SCHOOL MARKS

Test	School Marks	
	<u>r</u>	P.E.
Rate of Reading	.122 $\pm$	.074
Vocabulary	.578 $\pm$	.050
Comprehension, Part I	.338 $\pm$	.067
Comprehension, Part II	.458 $\pm$	.060
Comprehension Total	.446 $\pm$	.053
Composite Score	.561 $\pm$	.052
Monroe Rate Score	.139 $\pm$	.074
Monroe Comprehension Score	.280 $\pm$	.070
Otis I.Q.	.490 $\pm$	.057

The correlations between scores on all of the other tests and school marks may be regarded as significant, since the coefficients are at least four times the probable error. The correlation of school marks and the comprehension test devised in this study is higher than that of school marks and the Monroe test of comprehension. Both the vocabulary test and the composite score on the writer's test have higher correlation with school marks than has the Otis Self-Administering Test of Mental Ability, although the differences in size of coefficients are probably not significant.

In general, it appears that the tests which were made in this study are as highly correlated with school marks as such tests usually are. This fact does not necessarily mean that the battery furnishes a valid test of reading, as other factors, such as the correlation of both measures with a third variable, might be responsible for the correlation between the test battery and the criterion.¹ It is of considerable importance, however, that the

¹The correlation of composite scores on the new test and the Otis intelligence quotients of the Proviso pupils is .623. It might be thought that the correlation of intelligence with the results of this test and with school marks would furnish a satisfactory explanation of the correlation of the reading scores with school marks. Nevertheless, when intelligence was partialled out, and scores on the battery of tests were correlated with average school marks, the result was .372 $\pm$.065. As the correlation

criterion of agreement with school marks was applied to the test and that little or no evidence of low validity was secured.

The validity of the individual items in the test.- The validity of each item in the vocabulary and comprehension tests was studied by determining whether or not the pupils who made high scores on the test as a whole succeeded with the item in greater numbers than the pupils who made low scores on the test.¹ The method used was one which had been found to give results which agreed closely with biserial r ; it was utilized instead of biserial r because of the saving in time and labor involved.²

All the items in the tests showed positive validity by the method used. A few of the items in each part of the test had rather low validity. It is probable that the validity of the whole test, as measured by an outside criterion, might have been increased slightly by replacing the low-validity questions with more valid ones. It seemed unwise, however, to impair the equivalence of the forms of the test by such a procedure unless further research should indicate that the whole test was less valid than it appeared to be in this study.

Reliability of the Tests

Reliability of the rate test.- The correlations between the scores of several hundred ninth-grade public-school pupils on various forms of the test of rate of reading are shown in Table XV.

The data in Table XV support the finding reported in Chapter III that the reliability of the rate test is above .80. None of the correlations are below .80 and the mean of the three correlations is .825.

The interpretation of the reliability coefficients is complicated by the fact that there is no general agreement concerning what constitutes satisfactory reliability for tests. In fact, the question can be answered only in relative terms as the size of the correlation coefficient depends in considerable degree upon

coefficient is nearly six times the probable error, it is significant. It indicates that there is a relationship between performance on the test and success in school which is not explained by general intelligence.

¹It was impossible to use this procedure with the rate test.

²An explanation of the method used and the results secured are reported in the complete unpublished study which is filed in the library of the University of Chicago.

TABLE XV

THE RELIABILITY OF THE TEST OF RATE OF READING

Forms Correlated	Number of Cases	<u>r</u>	P.E.
1 and 2	309	.848 $\pm$	.011
3 and 4*	330	.823 $\pm$	.011
1 and 4	106	.804 $\pm$	.024

*A few changes, intended to increase the difficulty of the material, were made in Form 4, after the scores used in this correlation were obtained. However, the revision was so slight that it is improbable that it affected the relative standing of the pupils to any significant degree.

the variability of the scores used. Ruch says that when reliability coefficients are computed from the scores of a single class or grade, coefficients of .80 or .89 are "fairly high."¹ However, many widely used reading tests are not that reliable. Sangren states that most reliability coefficients which have been reported for standardized educational tests are between .60 and .80.² The present writer has found that statement to be substantially correct, both through a review of the literature concerning the reliability of reading tests and in his own study of standardized tests.

Reliability of the vocabulary test.- The correlation between the various forms of the vocabulary test are shown in Table XVI. The scores of 208 pupils in the ninth grade of the Parker Junior High School, Chicago, Illinois, were used in computing the correlation coefficients.

The correlations are quite high in view of the fact that each form contains only fifty items. The mean of the six coefficients, .851, is considerably higher than the correlation usually found for alternate forms of a vocabulary test of similar length. It was stated in Chapter III that a correlation of .667 $\pm$.04 was found for the forty-item word-meaning test which is a part of the Sangren-Woody Reading Test. It is probable that if the Inglis Test of English Vocabulary contained only fifty items, its reliability would be .75 to .80. This estimate is based on the reliability of .90+ which the publishers report for the entire test of 150 words.

¹Ruch, op. cit., p. 434.

²Sangren, op. cit., p. 32.

TABLE XVI

RELIABILITY OF THE VOCABULARY TEST

Forms	Number of Cases	<u>r</u>	P.E.
1 and 2	208	.860 $\pm$	.012
1 and 3	208	.872 $\pm$	.012
1 and 4	208	.840 $\pm$	.014
2 and 3	208	.852 $\pm$	.013
2 and 4	208	.836 $\pm$	.014
3 and 4	208	.847 $\pm$	.013

The mean probable error of a score on the forms of the vocabulary test is 2.28. This indicates that there is an even chance that the score a pupil made on the test is within 2.28 words of his true score. For instance, if a pupil made a score of 25, the chances are even that if he is given an infinite number of forms of the test, his mean score will be between 22.72 and 27.28. The probable error of a score is only about one-fourth of the standard deviation of the scores.

Reliability of the comprehension test.- Reliability coefficients for the total comprehension score, based on the scores of three groups of ninth-grade pupils are shown in Table XVII.

TABLE XVII

RELIABILITY OF THE TOTAL COMPREHENSION SCORE

Forms	Number of Cases	<u>r</u>	P.E.
1 and 2	150	.812 $\pm$	.019
2 and 3	54	.854 $\pm$	.026
1 and 4	53	.761 $\pm$	.040

The average of the three correlation coefficients shown in the foregoing table is .809. This barely falls within the range which Ruch has designated as high.¹ In general, the comprehension

¹Ruch, op. cit., p. 434.

test does not show quite as satisfactory reliability as do the rate and vocabulary tests.

Reliability of the composite score.- The reliability of the total scores on the test battery is shown in Table XVIII.

TABLE XVIII

RELIABILITY OF THE COMPOSITE SCORE

Forms	Number of Cases	<u>r</u>	P.E.
1 and 2	150	.908 <u>±</u>	.010
2 and 3	54	.950 <u>±</u>	.010
1 and 4	53	.912 <u>±</u>	.015

The table shows that the correlations between the various forms of the test battery are above .90. They fall within the category which Ruch has characterized as "high."¹

Only a few reading tests reach the reliability that is shown by this test. An extensive body of data on the reliability of comprehension tests has been reported by Current and Ruch² and by Mosher,³ who followed the technique of Current and Ruch with the same tests which they studied. These investigators studied the reliability of the following reading tests: Chapman, Courtis, Haggerty, Monroe, Thorndike-McCall, and Stanford Achievement. Of the six tests, the Stanford Achievement reading test was found to be the most reliable in both studies. The reliability coefficients reported for that test by Current and Ruch and by Mosher were .93 and .92, respectively. These coefficients are approximately the same as those found for the new battery of tests in the present study. It may be concluded, therefore, that the new test is quite consistent in measuring general reading ability, in comparison with other tests in the field.

¹Ruch, op. cit., p. 434.

²W. F. Current and G. M. Ruch, "Further Studies of the Reliability of Reading Tests," Journal of Educational Psychology, XVII (1926), 476-81.

³R. M. Mosher, "A Further Note on the Reliability of Reading Tests," Journal of Educational Psychology, XIX (1928), 272-74.

The Utility of the Tests

It was stated in Chapter II that the main purpose for which the tests were constructed was to furnish a means of making a number of rapid, valid, and fairly comprehensive measurements of general reading skill in order that a unified program of testing might be carried on over an indefinite period in the junior high school. The internal analysis of the tests that has been made in this chapter indicates that they fulfill this purpose. It has been shown that the tests are workable, objective measuring instruments, that they contain four equivalent forms, that they are apparently valid for the purposes intended, and that each of the three major parts is reliable enough to be used in individual diagnostic work. The utility of the tests in meeting the major objective for which they were designed will be demonstrated in the next chapter.

The tests lend themselves, furthermore, to certain subsidiary uses, because of the number of duplicate forms. For example, one may wish to make exceptionally reliable measurements of certain aspects of reading ability. This can be done by making use of more than one form of that part of the reading test which is applicable to the skill which is being considered. For instance, if one wished to find, with great accuracy, how rapidly a group of children read, he might use all four forms of the rate test. These forms, including the comprehension questions based on the rate material can be administered in less than an hour. The reliability of the various parts of the test, when one to four forms are used consecutively, is shown in Table XIX. The Spearman-Brown

TABLE XIX

THE RELIABILITY OF THE READING TESTS WHEN SEVERAL FORMS
ARE USED CONSECUTIVELY

Test	One Form	Two Forms	Three Forms	Four Forms
Rate	.825	.904	.934	.950
Vocabulary	.851	.919	.945	.958
Comprehension of continuous material . . .	.648	.787	.848	.880
Power of comprehension	.727	.842	.887	.914
Total comprehension	.809	.894	.927	.944
Composite score	.923	.951	.973	.977

prophecy formula for lengthened tests was used in finding the

reliability of more than one form.

Table XIX is read as follows: The reliability of a single form of the test of rate of reading is .825; the reliability of two forms used together is .904; the reliability of three forms is .934; and the reliability of four forms is .950. The rest of the table is read in similar fashion.

This table shows that it is possible to secure a highly reliable measurement of various aspects of reading ability by using several forms of the tests together. The capacity of the tests for both extensive and intensive measurement probably makes them exceptionally serviceable for testing the reading ability of junior-high-school pupils.

CHAPTER V

EXPERIMENTAL TEACHING OF CORRECTIVE READING¹

The Problem of the Chapter

The work of this investigation up to the present point has been preliminary to the most important problem. It has been reported in considerable detail because it has dealt with the development of adequate measuring instruments, which are basic to successful scientific experimentation with teaching procedures. An attack will now be made upon the central problem of the study, namely, the improvement of the silent reading achievement of retarded readers at the junior-high-school level when they are taught as a group under regular classroom conditions.

The specific purposes of this part of the investigation, as stated in Chapter II, are as follows:

1. To study experimentally the value of teaching corrective reading to a group of seventh-grade pupils.
2. To develop techniques for use in correcting the reading defects of seventh-grade pupils.
3. To investigate the permanence of gains made as a result of corrective teaching.

The main part of the chapter presents a description and evaluation of experiments in corrective reading which were carried on in the seventh grade of the University High School of the University of Chicago during the years 1929-30 and 1930-31.

The Initial Experiment

The classes used.— The first controlled experiment in corrective reading in the University High School was conducted during the second semester of the year 1929-30.² Reading test scores and

¹For a definition of corrective reading, as the term is used in this study, see page 12.

²There were two earlier experiments which had no connection with the present study. For an account of them, see The English Department of the University High School, "Differentiating Instruction in Ninth Grade English," School Review, XXVII (December, 1919), 772-88; and Edith E. Shepherd, "Some Silent Reading Lessons in Junior High School English," School Review, XXIX (March, 1921), 206-15.

teachers' reports secured during the first semester indicated that eleven seventh-grade pupils and eleven ninth-grade pupils were much below standard in reading and that a considerable number of others were not strong in either reading or class work. It happened that most of the inferior readers were concentrated in a ninth-grade class in Community Life-English and in a seventh-grade English class. Some of the children in these classes were up to or above standard in reading, but the majority of them were below average. As resectioning the classes on the basis of reading ability would have been impracticable so late in the year, it was decided to utilize these two classes for purposes of an experiment in teaching corrective reading to groups of pupils.

The plan of the initial experiment.- The classes were taught by the regular classroom teachers, each of whom was aided by a research assistant.¹

The reading in the ninth grade was incidental to the regular work of the course. The Community Life content was emphasized and attention given to reading only at points where the pupils encountered difficulties because of deficiencies in reading. The experiment in this class lasted from January 7 until the end of the school year in June.

The experiment in the seventh grade was conducted for only six weeks--January 7 to February 14. The attack upon reading was much more direct and intensive than it was in the Freshman class. The pupils were made conscious of their deficiencies and were led to understand that the major purpose of the class during those weeks was to bring their reading up to standard. Class materials were selected for their reading value rather than for their subject-matter value. Tests were given frequently and the pupils were told where they stood. Every effort was made to motivate the pupils to put forth their best work to improve their reading.

Results of the experiment.- A seventh-grade English class in which no attention was given to reading, other than the incidental attention that came about in the regular work of the course, was used as a control group for both experimental classes. Form 2 of the Monroe Standardized Silent Reading Test III was given to all three classes at the beginning of the period of corrective teaching. At the end of the seventh-grade experiment in February,

¹Miss Edith Shepherd, Chairman of the English Department of the University High School, taught the ninth-grade class and Miss Helen Lacey, an instructor in the department, taught the seventh-grade class. Ivan A. Booker assisted Miss Shepherd; the writer assisted Miss Lacey.

Form 1 of the Monroe test was administered to that class and to the control group. The comparative standing of the two groups is shown in the following table.¹

TABLE XX

GAINS IN READING MADE BY THE EXPERIMENTAL GROUP AND THE
CONTROL GROUP IN THE PRELIMINARY EXPERIMENT IN THE
SEVENTH GRADE

	Experi- mental Group	Control Group
Number of pupils in each group	27	26
Number of pupils who gained in rate during the training period	24	16
Number of pupils who lost in rate during the training period	3	10
Number of pupils who gained in comprehension during the training period	17	13
Number of pupils who lost in comprehension dur- ing the training period	7	11
Number of pupils who neither gained nor lost in comprehension	3	2
Number of pupils who gained in both rate and comprehension	15	10
Number of pupils who lost in both rate and com- prehension	1	4
Average gain in rate	19.6	6.6
Average gain in comprehension	6.0	1.9

The entries in Table XX furnish definite evidence that the reading instruction was effective. The gains made by the remedial group were superior to those made by the control group in every instance. Particular attention is called to the fact that fifteen children in the corrective group exhibited gains in both rate and comprehension, while only ten pupils in the control group showed gains in both of these scores. The average gain in rate made by the pupils in the corrective group was thirteen points more than

¹"Remedial Reading," op. cit., p. 908.

that made by those in the control section, while their gain in comprehension was 4.1 points more than that of the control group. In view of the brevity of the training period, these gains are quite remarkable. Although it was shown in Chapter II of this study that the Monroe test is not thoroughly satisfactory from the standpoint of reliability, such decidedly favorable results on the test indicate that considerable gain did, in fact, take place.

At the end of the ninth-grade experiment scores on a number of tests, when compared with scores on entrance tests in reading which the pupils in both the experimental and the control group had taken, indicated that the experimental class had made greater gains in reading than had the control class.

In general, the preliminary experiment in teaching backward readers at the junior-high-school level in groups yielded results that were sufficiently positive to warrant a more extensive experiment the following year.

The Experiment in 1930-31

The reading experiment during the year 1930-31 was carried on in the seventh grade of the University High School. It comprised the following major steps:

1. Testing the reading ability of all pupils in the seventh grade.
2. Analysis of data and the selection of retarded readers for corrective teaching.
3. Selection of the control groups.
4. Diagnosis of the difficulties of the pupils in the corrective-reading group.
5. Planning teaching procedures to meet the needs of the pupils.
6. Doing the teaching.¹
7. Testing the pupils for improvement in reading during and at the end of the period of training.
8. Testing the pupils for permanence of gains some months after the end of the teaching.
9. Studying the findings and drawing conclusions about the results of the experiment.

¹This part of the experiment was conducted by Miss Helen Lacey, the instructor who had taught the seventh-grade class during the preliminary experiment. The planning of the teaching was done by Miss Lacey and the writer. The rest of the steps were carried on mainly by the writer.

Testing Reading Ability

A battery of tests was administered in May, 1930, to the pupils in the sixth grade of the University Elementary School and to pupils from other schools who were to enter the seventh grade of the University High School the following autumn. The battery included the following tests: the Otis Self-Administering Test of Mental Ability, Higher Examination; rate and comprehension scores on the Monroe Standardized Silent Reading Test, Test III; the Thorndike-McCall Reading Scale, the Inglis Test of English Vocabulary; and two tests devised by the writer which, for purposes of identification in the discussion of test results, will be called the U. H. S. Test of Rate of Continuous Silent Reading and the U. H. S. Reading Vocabulary Test.¹

Analysis of Test Scores and Selection of Reading Cases

A weakness in the first experiment.- In the preliminary study, the pupils in the reading groups were classified on the basis of ability in English usage, rather than on the basis of reading ability. This was due to the fact that the term had started before the results of the tests were analyzed and the poor readers identified. Consequently, a considerable number of the students in the experimental classes did not need special attention in reading, while others who did need the training but were not in the classes had to be taught individually. That difficulty was avoided in the experiment of 1930-31 through the following procedure in selecting the cases.

Classifying the poor readers.- The results of the tests were studied carefully, soon after they were given, for evidence of weakness in reading. A list was made of the pupils who were in the lower quarter of their class on two or more reading tests.² The list was submitted to the pupils' sixth-grade teachers in the Elementary School for criticism and comparison with the standing of the pupils in their class work. Then the test results were studied again in the light of the suggestions made by the teachers,

¹The U. H. S. tests included in this entrance battery were not identical with the U. H. S. tests which were used later to measure gains in reading. The rate test was the one which dealt with Indian games, mentioned near the beginning of Chapter III, and the vocabulary test contained 100 items instead of 50, the number used later.

²The Otis Self-Administering Test of Mental Ability was regarded as a reading test during the making of the list.

and a revised list was drawn up. Very few changes were necessary as the agreement between test scores and teachers' judgment was very close. The revised list was taken to the Head of the English Department of the University High School who scheduled the children whose names appeared on it for an English class in which special attention was to be devoted to reading.

Selecting the Control Groups

Two seventh-grade English classes, taught by the same instructor who did the teaching of the experimental group, were selected as control groups. These classes contained pupils who were average or superior in both general ability and reading skill. It would, of course, have been desirable to have secured a control group composed of pupils whose intelligence and initial reading ability were equivalent to that of the experimental pupils. This, however, was impossible, since all of the seventh-grade pupils who were shown by the entrance tests to be retarded readers were placed in the experimental group.

The fact that it was not possible to equate the control groups with the experimental pupils is a limitation of the study, as there is no certainty that a control group composed of retarded readers would have made the same amount of progress on the tests as did the control groups made up of average and superior readers. This limiting factor was kept in mind in interpreting the results.

Diagnosing the Difficulties

Summary of test scores.- Table XXI shows the median scores of the reading group in comparison with the test norms and the class medians on the entrance tests.

A study of Table XXI and of the reading scores made by each individual in the class indicated that the status of the class in corrective reading at the beginning of the experiment was as follows:

1. The pupils were typical of an unselected group of school children in general intelligence. The median Otis I.Q. was 99 and the range was from 85 to 112. However, the median I.Q. of this group was 12.2 points below the median I.Q. of the entire seventh grade of the University High School. Only two members of this group reached the median of the whole class in intelligence.

2. Most of the pupils were below norm in reading rate. This is shown by both the Monroe and the U. H. S. tests.¹ Only

¹The norms for the U. H. S. test were only tentative since the test was not fully standardized.

TABLE XXI

SCORES OF CORRECTIVE GROUP ON ENTRANCE TESTS IN READING

Test	Reading Group Median	Test Norm	U. H. S. 7th Grade Median
Otis Self-Administering (I.Q.)	99.0	100.0	112.2
U. H. S. Rate of Continuous Reading . .	2.75	3.25	3.34
Monroe Rate	77.0	83.0	106.0
Monroe Comprehension	25.5	23.0	35.5
Thorndike-McCall (T-score)	56.7	58.3	63.8
Inglis Vocabulary	26.7	--	39.5
U. H. S. Reading Vocabulary	21.7	40.0	41.2
Standardized Oral* (weighted score) . .	40.3	48.0	--

*The Gray Standardized Oral Reading Paragraphs were not a part of the battery of tests given in May. They were administered to each of the pupils in the reading group soon after the opening of the fall term.

three pupils were above the norms on both rate tests. In comparison with the median of the seventh grade of the University High School, the children in the reading class read .59 words per second slower on the U. H. S. test and were 29 points slower on the Monroe test.

3. The medians of the comprehension scores on the Monroe test and the Thorndike-McCall test were not far from the norms of the tests. However, certain pupils were decidedly low in comprehension. The class as a whole was 10 points below the median of the entire grade on the Monroe test and 7.1 points below it on the Thorndike-McCall Reading Scale.

4. The class was very low in vocabulary according to the results of the Inglis test and the U. H. S. test. All of the pupils except three appeared to be below standard in knowledge of word meaning. The median of the group was 12.8 points below the median of the seventh grade on the Inglis test and was 19.5 points below it on the U. H. S. test.

5. The median weighted score of the group on the Standardized Oral Reading Paragraphs was 7.1 points below the eighth-grade norm. Only two pupils exceeded the norm. Unquestionably, the class was composed of inferior oral readers.

The composite picture of the reading class at the beginning

of the year was that of a group of pupils average in intelligence but retarded in reading. This retardation was especially marked in rate and in knowledge of word meaning. In comparison with the other seventh-grade pupils of the University High School, the class was decidedly low at every point measured by the battery of tests.

The most important reading needs of the class.- The test scores supplied but one general type of information about the reading difficulties of the pupils. While they furnished a starting point for individual diagnosis and made possible such statements as have already been made concerning the reading ability of each pupil, the tests did not supply enough data to form the sole basis of the reading program.

Accordingly, the pupils were observed carefully by the teacher and the research worker during their English class for a period of two weeks before the experiment was launched. Notes were made about the reading interests of the pupils, the kinds of deficiencies which they manifested, and the points where failure to understand the content of the unit seemed to be due to poor reading. These observations, added to the information gained by test scores, led the teacher and the research worker to classify the major reading needs of the group under the following general heads:

1. A need to form genuine reading interests and to do much independent reading. In spite of the competent guidance in reading activities which most of them had received in the elementary school, many of the children had acquired little interest in reading. One of the most important tasks which the teachers faced was to develop real reading interests and to see that the pupils were surrounded with desirable reading material.

2. A need to learn to do certain necessary types of reading. It was quite evident in the very beginning of the work with this class that the children did not know how to read with a purpose in view. They could do a certain kind of narrative reading fairly well, but they had little conception of how to attack descriptive or expository material so as to get an essential understanding of what it contained. In other words, the pupils did not know how to study. It became a major responsibility of those in charge of the class to teach the pupils how to read in a study situation so that they might be able to function successfully in high school.

3. A need to acquire skill in fundamental reading habits. Lack of adequate mastery of reading habits was shown both in the test results and in the class work of the pupils.

4. A need to acquire a wider reading vocabulary. It was apparent in the entrance test scores that twelve of the pupils were very deficient in reading vocabulary. Obviously, they could not

read effectively in any material except that which was highly juvenile in nature until they had gained a wider knowledge of the meaning of words.

5. A need to gain power of interpreting the meaning of sentences, paragraphs, and whole selections. The pupils needed to learn to think as they read and to assimilate and relate the reading material to their past experience.

Planning the Teaching Procedures to Meet the Reading Needs

The next step was to decide upon the kinds of activities in which the pupils should engage to meet their reading needs. It was, of course, impossible to produce a ready-made plan at the beginning and to carry it through, unaltered, step by step. Inevitably, the teachers found it necessary at times to resort to a kind of trial-and-error procedure—to make the best possible guess concerning the most desirable way of meeting the reading difficulties of the pupils—and then to experiment and readjust the teaching until progress was secured.

However, before launching the corrective program, the experimental literature in the field was canvassed for suggestions about procedures for handling the kinds of reading needs which the diagnosis of difficulties had revealed. This led to the adoption, with modifications at certain stages of the experiment, of some of the procedures used with success in previous experiments.

The organization of the class work.—The various types of reading activity used in the experiment were organized into the following units:

1. Developing genuine reading interests and forming habits of regular independent reading.
A part of each period was devoted to silent reading with a minimum of supervision.
2. Learning to read and follow directions.
 - (1) Printed directions preceding the reading.
 - (2) Directions written on the blackboard.This skill was emphasized throughout every reading unit and also in the units in grammar. Almost every lesson had as a secondary aim reading to follow directions, and this was made a primary aim when necessary.
3. Learning to find facts in a paragraph.
 - (1) Reading to find certain definite facts.
 - (2) Reading and then answering fact questions from memory of the paragraph.
4. Learning to find main ideas
 - (1) in single paragraphs.

- (2) in a series of paragraphs organized about one central idea.
 - (3) in a whole selection.
 - 5. Learning to find details which support main ideas in paragraphs and to recognize them as supporting material, when found
 - (1) near the beginning of a paragraph.
 - (2) near the end of a paragraph.
 - (3) throughout a paragraph.
 - (4) in succeeding paragraphs.
 - 6. Learning to find certain definite points through rapid reading of a selection covering several pages.
 - (1) Words, such as names of animals, games, or inventions.
 - (2) Ideas, expressed in phrases, sentences, or paragraphs.
- This was the first point in the training where the attention of the children was directed toward their rate of reading. Comprehension remained uppermost, but the reading was timed and the children were made aware of their varying rates of reading.
- 7. Learning to identify the large general divisions or links in longer selections.
 - (1) Narrative reading.
 - (2) Descriptive reading.
 - (3) Expository reading.
 - 8. Building up power of grasping and interpreting meaning through
 - (1) oral reading of literature in small groups and later telling the stories read to the rest of the class.
 - (2) silent reading to answer thought questions.
 - (3) silent reading and the discussion of questions which require interpretation of what has been read.
 - (4) reading under the pressure of test conditions.
 - (5) studying new words.

The sources of reading materials.- Three plans for securing suitable reading material were tried in this experiment. During the first part of the year, the content of the regular course in Community Life-English formed the basis of the reading instruction. The textbook in the course was Hill's Community Civics.¹ However,

¹Howard C. Hill, Community Civics. Boston: Ginn and Company, 1928.

the content of the course was exceedingly rich and varied, and the textbook was but one of many sources of material.

Under this organization of the work in reading, the teacher was obliged to keep two general objectives in mind: (1) to get the content across to the pupils, and (2) to help the pupils develop certain specific skills in reading. It was discovered that the finding of material that would serve the purposes of a Community Life unit, and would, at the same time, give training in a particular type of reading was sometimes exceedingly difficult and often consumed a very large amount of the teacher's time.

In the second part of the course, comprising Units V, VI, and VII, the reading matter was selected with a view to getting the material which would best promote the reading skill in which practice was being given. No effort was made to organize the reading around any definite body of content. The main sources of material were Lyman and Hill's Literature and Living, Books I and II,¹ and Hill's Readings in Community Life.²

Near the end of the experiment, the plan of basing reading instruction upon modern literature materials was tried. The aim of the teaching was divided between appreciation of the selections read and development of reading skill. This duality of purpose was somewhat analogous to that which obtained at the beginning of the experiment, but the organization of the reading materials was characterized by greater mobility than was that of the Community Life course.³

The distribution of time used in the experiment.- In all, eighteen weeks of four periods each, or sixty-nine days, excluding

¹Rollo L. Lyman and Howard C. Hill, Literature and Living. New York: Charles Scribner's Sons, 1925.

²Howard C. Hill, Readings in Community Life. Boston: Ginn and Company, 1930.

³No objective evaluation of the different sources of reading matter was attempted in the experiment. The specific plan adopted for getting reading materials was not directly related to the problem of the study, inasmuch as all three plans enabled the teachers to work specifically upon techniques. In general, it was found that it was easier to organize the teaching of reading when the subject matter was drawn from a variety of fields. The main objection to this plan for getting reading materials was that the pupils tended to become misfits administratively in that, by the end of the year, they lacked certain fundamental understandings which the other children in their grade had gained by studying particular kinds of subject matter.

holidays, were devoted to the experiment. The time was distributed as follows: five weeks in the first semester (October 14 to November 14); nine weeks near the beginning of the second semester (February 10 to April 17, leaving out the week of March 23, in which spring vacation occurred); and four weeks at the end of the second semester, starting on May 19. The time between these periods of reading instruction was given over to certain units in grammar.

The testing program.- An exceptionally extensive program of testing with standardized and non-standardized reading materials was planned and carried on throughout the experiment. The Sangren-Woody Reading Test and the Standardized Oral Reading Paragraphs were administered at the beginning and the end of the experiment. Forms of the tests devised in the first part of this study, which were not ready for use until the beginning of the second semester, were used before and after the second period of reading instruction, at the end of the third period of instruction, and in October, 1931, several months after the experiment was brought to an end. The Monroe Standardized Silent Reading Test, Test III, was used at the beginning and the end of the second period of instruction. The Otis Self-Administering Test of Mental Ability was administered at the end of the experiment. Various special diagnostic techniques, such as the apparatus for photographing eye-movements and a test of rate of association were used with each pupil. Numerous informal tests devised by the teacher of the class or by the research worker were aids to instruction at every stage of the experiment.¹ The results of the testing procedures will be discussed in detail near the end of the chapter.²

Plan of recording the work of the experiment.- The research worker kept a daily record of the activities of the class. This record was made in the form of notes during the class hour and was rewritten after class. It was utilized extensively by the teacher and the research worker in working out the lesson plans for succeeding days.

The Teaching

No attempt will be made to present a detailed report on the teaching procedure used in this account of the work of the

¹In the account of the experiment, Miss Helen Lacey, the instructor of the class, is referred to as the teacher; the writer is called the research worker; and Miss Lacey and the writer, taken collectively, are termed the teachers.

²Pages 70-79.

class. A full record of the minutes of the class meetings is in an unpublished report.¹

The number of class periods that were spent on each of the units was conditioned by the speed with which the pupils learned to do the type of reading that was emphasized in it. When the teacher was ready to undertake a new unit with the class, she made out and administered a pretest, which was designed to find out how much reading skill of that particular kind the pupils already possessed. If the scores were low, she trained the pupils during several class periods until her observation, the notes taken by the research assistant, and informal tests given by him, indicated that the pupils had made significant progress. Then the pretest was repeated. If most of the pupils made scores on it that were very high, the class took up a new unit. If not, they continued to work on the old one until a satisfactory level of attainment was reached.

Although the pupils in the class were taught for the most part as a group rather than individually, the needs of individual pupils were taken into account in the procedure. Whenever reading scores and observation by the teachers indicated that certain pupils did not need further training in the particular skill on which the class was working, they were not required to participate in the work of the class but were allowed to do free reading.

A few of the pupils who were especially handicapped in certain phases of reading met with the research worker from time to time. For example, one pupil who made a very poor record when his eye movements were photographed spent two periods a week for five weeks receiving special help from the research worker. Flash card exercises and the reading of interesting stories under timed conditions formed the basis of the work in these periods.

Several other pupils met with the research worker a number of times for special study of the meaning of words. However, these individual conferences were not carried beyond the limits that a classroom teacher might have taken them, as the purpose of the experiment was to make a study of procedures which would be applicable to the usual conditions of a junior high school.

Results of the Experiment

Six sources of information were used in studying the gains that were made by the pupils in the reading class. They were as follows:

¹Report of the Research Workers in Reading for 1930-31 in the files of the Research Secretary of the University Laboratory Schools.

1. Scores on the Sangren-Woody Reading Test.
2. Scores on the Monroe Standardized Silent Reading Test.
3. Scores on the U. H. S. Silent Reading Test.¹
4. Scores on the Gray Standardized Oral Reading Paragraphs.
5. Eye-movement records of some of the pupils.
6. Changes in I.Q. as revealed by the Otis Self-Administering Test of Mental Ability.

Not all of the results will be presented in detail. The discussion will be confined to the scores on the Sangren-Woody Reading Test, the scores on the U. H. S. test, a summary of the gains on all the tests, and a summary of the reading status of the individual pupils at the beginning and the end of the experiment.

Scores on the Sangren-Woody Reading Test.- The Sangren-Woody Reading Test was used to measure the gains in silent reading made by the class between the time of beginning and the time of ending the experiment. This test, which has two forms, was described briefly in Chapter II.

Form A of the Sangren-Woody test was administered to the reading class on October 15, 1930, and Form B was administered on June 11, 1931. These forms were given to a control group on October 15 and June 15, respectively. The control group consisted of twenty-nine pupils in one of the other seventh-grade English classes.

The means of the scores that were made on the different parts of the Sangren-Woody test by the pupils in the reading class are compared with the means of the scores that were made by the control-group pupils in Table XXII.

This table shows that the experimental group made a substantial gain on six of the seven parts of the test. The gain in total score is 14.0, which is 17.3 per cent of the mean total score made by the group in October. The gains of the experimental group exceeded those of the control group in every part of the test. The loss in total score shown by the control group is probably due to a slight difference in the difficulty of the forms of the test.²

Since the experimental and control groups were not equated in initial reading ability, a question might be raised concerning whether or not the experimental group gained more than the normal

¹The U. H. S. Silent Reading Test is the test constructed in this study.

²Scores for this test that are in the files of the University High School indicate that Form B is slightly harder than Form A. The largest differences appear to be in word meaning and following directions.

TABLE XXII

MEAN SCORES MADE BY THE PUPILS IN THE CORRECTIVE READING CLASS AND IN THE CONTROL GROUP ON THE VARIOUS PARTS OF THE SANGREN-WOODY READING TEST

Part	Experimental Group			Control Group			Superiority of Exp. Group in Gain
	Oct.	June	Gain*	Oct.	June	Gain*	
Word Meaning	23.0	27.0	4.0	29.0	27.3	-1.7	5.7
Rate of Reading . .	19.2	24.4	5.2	23.7	27.4	3.7	1.5
Fact Material . . .	9.4	11.0	1.6	13.8	12.1	-1.7	3.3
Total Meaning . . .	6.0	7.6	1.6	8.9	8.4	-0.5	2.1
Central Thought . .	7.0	8.4	1.4	9.3	9.2	-0.1	1.5
Following Directions	7.3	7.0	-0.3	9.1	7.5	-1.6	1.3
Organization	9.1	9.6	0.5	10.5	10.7	0.2	0.3
Total Score	81.0	95.0	14.0	104.3	102.6	-1.7	15.7

*A minus sign (-) preceding a number in the gain column indicates a loss in mean score.

amount for a group at its level of reading ability. A comparison of the initial and final scores of the experimental group with the norms for the Sangren-Woody test indicates that this question can be answered affirmatively. The class was brought, in a period of eight months during which approximately half of the time was devoted to reading, from a position three points below the seventh-grade norm in mean total score to a place one point above the eighth-grade norm. The average gain in reading score was 1.4 grades. At the end of the year, the mean scores of the class were up to or above the eighth-grade norm on all parts of the test except fact material and total meaning.

Scores on the U. H. S. Reading Test.- This test was not administered at the beginning of the experiment, as it was in the process of construction at that time.¹ Form 1 was given at the beginning of the second period of instruction in reading, in February, and Form 2 was used at the end of that period in April. Form 3 was administered at the end of the year in June, and Form 4 was given in October of the following school year.

¹See Chapters III and IV for a detailed description and evaluation of the test.

Twenty-nine pupils in one of the other seventh-grade English classes were used as a control group for this test. This was not the same control group that was used with the Sangren-Woody Reading Test.

The means of the scores which the pupils in the experimental and control groups made on the first three forms of the U. H. S. test are shown in Table XXIII.

TABLE XXIII

MEAN SCORES MADE BY THE EXPERIMENTAL GROUP AND THE CONTROL GROUP ON FORMS ONE, TWO, AND THREE OF THE U. H. S. READING TEST, ADMINISTERED ON FEBRUARY 10, APRIL 21, AND JUNE 12, RESPECTIVELY

Test	Experimental Group				Control Group				Superiority* of Exp. Group in Gain
	Form 1	Form 2	Form 3	Gain* F. 1 to F. 3	Form 1	Form 2	Form 3	Gain* F. 1 to F. 3	
Vocabulary . .	17.8	18.4	20.1	2.3	28.8	30.8	31.4	2.6	-0.3
Rate	33.5	35.7	37.9	4.4	42.0	39.5	41.7	-0.3	4.7
Comprehension	31.5	33.3	35.3	3.8	43.2	45.2	43.7	0.5	3.3
Total Score .	82.8	87.4	93.3	10.5	114.0	115.5	116.8	2.8	7.7

*A minus sign (-) in the "gain" column indicates a loss in score; a minus sign in the column headed "Superiority of experimental group in gain" indicates superiority on the part of the control group.

The means of the scores of the experimental group show a steady gain on all parts of the test. The means of the scores of the control group fail to show a similar consistent gain, although the control group made much higher average scores than did the experimental group, as it consisted, for the most part, of readers of superior ability.

In rate and comprehension, the experimental class made a considerably larger gain from Form 1 to Form 3 than did the control group. The control pupils, however, showed a slightly larger gain than the experimental group in vocabulary, although their superiority in that respect was insignificant. The fact that the control group made a considerable increase in vocabulary score suggests that the word-meaning part of Form 3 may be a little easier than

the corresponding part of Form 1. It has already been shown, however, that the scores of 208 public-school pupils indicate that there is no significant difference in the difficulty of the various forms of the vocabulary test.¹ It is possible that some undiscovered incidental factors in the regular class work of the control group helped to increase the vocabulary of the pupils in that class.

The experimental group gained 10.5 points in mean total score on the test, while the control group gained only 2.8 points in mean total score. The gain of the experimental group is unquestionably significant as it is approximately 6.8 times the probable error of the difference between the means of the scores on the first and third forms; the gain made by the control group does not appear to be significant. Since it was shown in Chapter IV that the forms of the test are practically equivalent, the foregoing data furnish excellent evidence that the experimental class, in general, improved in reading ability between February 10 and June 12.

Summary of test results.- The gains of the reading class in scores on the various reading tests which were used with it during the experiment are summarized in Table XXIV.

TABLE XXIV

MEAN GAINS IN TEST SCORES MADE BY THE READING CLASS

Test	First Adminis-tered	Last Adminis-tered	Length of Period	Points Gain in Score	Per Cent Gain in Score
Sangren-Woody . . .	October	June	8 months	14.00	17.3
U. H. S.	February	June	4 months	10.50	12.7
Monroe Rate	February	April	2 months	18.91	21.9
Monroe Compr. . . .	February	April	2 months	4.53	18.8
Oral Reading	October	June	8 months	4.50	11.4
Vocabulary	May, 1930	June, 1931	13 months	16.50	71.1
Otis S-A I.Q.	October	June	8 months	2.10	2.1

The preceding table offers ample evidence that the instruction given to the corrective reading class during the training period was effective. The experimental group made gains in reading ability according to every test which was given. The control

¹See page 39.

group, on the other hand, lost 1.7 points or 1.6 per cent on the Sangren-Woody test and gained only 2.8 points, or 2.5 per cent on the U. H. S. test.¹

A summary of the reading status of individual pupils at the beginning and the end of the experiment.- The number and percentage of pupils in the experimental class who were below the seventh-grade norms on the first and last administrations of the Sangren-Woody Reading Test, the U. H. S. Silent Reading Test, and the Standardized Oral Reading Paragraphs are shown in Table XXV.

TABLE XXV

NUMBER AND PERCENTAGE OF PUPILS IN EXPERIMENTAL GROUP BELOW NORM WHEN SANGREN-WOODY, GRAY, AND U. H. S. TESTS WERE FIRST ADMINISTERED COMPARED WITH NUMBER AND PERCENTAGE BELOW NORM WHEN THE TESTS WERE GIVEN AT THE END OF THE YEAR

Test	Norm	Number below Norm 1st Admin.	Number below Norm 2nd Admin.	Decrease	Percent- age below Norm 1st Admin.	Percent- age below Norm 2nd Admin.	Decrease
Sangren-Woody							
Word Meaning . .	25	17	6	11	70.8	25.0	45.8
Rate	21	16	8	8	66.7	33.3	33.3
Comprehension* .	38	11	4	7	45.8	16.7	29.1
U. H. S.#							
Vocabulary . . .	20	16	9	7	66.7	37.5	29.2
Rate	32	11	5	6	45.8	20.8	25.0
Comprehension .	30	10	1	9	41.7	4.2	37.5
Gray Oral (Raw score)	34	16	8	8	66.7	33.3	33.3
Mean		13.9	5.9	8.0	57.7	24.4	33.3

*The scores on the last five parts of the Sangren-Woody Reading Test were combined for a total comprehension score.

#These norms for the U. H. S. test were made in a rather arbitrary fashion by taking into consideration the performance of pupils in public schools and in the University High School. More recent experimentation indicates that they are too high for unselected seventh-grade children.

¹The other tests were not administered to the control group.

The decrease in percentage of pupils below the seventh-grade norm varied from 25 per cent on the U. H. S. rate test to 45.8 per cent on the Sangren-Woody word meaning test. On the average, one-third of the pupils in the class raised their scores from below norm to above norm. Less than 25 per cent of the scores on the three tests were below norm at the end of the year.

The Permanence of the Gains in Reading

Previous studies in corrective reading have, with few exceptions, neglected the question of whether or not the gains made during corrective teaching were permanent ones. As conclusions about the results of corrective work in reading have more value if the answer to this question is known, the seventh-grade experimental classes used in this study were retested several months after training was discontinued.

The permanence of the reading gains in the first experiment.- In October, 1930, approximately seven months after the initial experiment was brought to a close, the Monroe Standardized Silent Reading Test, which had been used to measure gains during the training period, was administered again to the pupils who had been in the experimental class. Of the twenty-four pupils who took the test in October, sixteen or 67 per cent were up to or above their rate of reading on the February test, and all but four of the pupils exceeded their rate on the January test. In comprehension, thirteen or 54 per cent were up to or above their February score, while all but four of the children surpassed their January score. All except two of the pupils came up to or went above their February performance either in rate or in comprehension.

The mean rate score made by the group on the October test surpassed the February score in rate of reading by 2.3 points, and the mean comprehension score was exactly equal to the February score. Both of the October means were much above the January means.

The preceding facts indicate that considerable permanent benefit was derived from the six weeks of corrective instruction in reading. This general conclusion is, however, limited by two factors: (1) more than half of the period between the end of the teaching and the retest included time when the children were in school and were probably making some improvement in reading as a result of incidental classroom training; and (2) as the Monroe test has only two forms, the form administered in October had been taken by the pupils during the experiment, and specific practice may have affected the results. These two limiting factors were eliminated from the retest that followed the main experiment, carried on in 1930-31.

The permanence of the gains in the second experiment.- It

was stated in a preceding section of this chapter that Forms 1, 2, and 3 of the U. H. S. Silent Reading Test, which contains four equivalent forms, were administered to the experimental and control groups during the second and third periods of corrective teaching. In order to find out whether the gains made during the experiment were retained, Form 4 was administered in October, 1931, approximately four months after the experiment had ended.

The period between the discontinuance of teaching and the test in October covered the summer vacation. The results of research studies indicate that, in general, pupils usually make little, if any, gain in reading during a summer vacation.¹

The test was administered to three Freshman English classes at one of their study periods. These classes contained, in addition to a number of other pupils, twenty-two pupils who had been in the experimental class and twenty-five pupils who had been in the control group. There was nothing in the procedure of giving the test to suggest to the experimental pupils that the test had any connection with the class in corrective reading.

The mean scores that were made by the experimental group and the control group in June and October are shown in Table XXVI.

This table shows that the experimental group made a very slight gain from Form 3 to Form 4 in vocabulary and rate, but showed a small loss in comprehension. The loss in total score was 0.65 point. None of the differences between the mean score on Form 3 and the mean score on Form 4 are significant differences. In general, the evidence indicates that the experimental group neither gained nor lost in reading ability during the period following the work in corrective reading.

The control group gained in rate of reading, but lost in

¹Elizabeth Bruene, "Effect of the Summer Vacation on the Achievement of Pupils in the Fourth, Fifth and Sixth Grades," Journal of Educational Research, XVIII (November, 1928), 309-14.

M. J. Nelson, "The Abilities and Achievements of Elementary School Pupils before and after a Vacation," School and Society, XXVIII (September, 1928), 371-72.

J. C. Morrison, "What Effect Has the Summer Vacation on Children's Learning Ability and Ability to Learn?" Educational Research Bulletin, III (October 1, 1924), 245-49. Columbus, Ohio: Bureau of Educational Research, Ohio State University.

Margaret E. Noonan, Influence of the Summer Vacation on the Abilities of Fifth and Sixth Grade Children. Teachers College Contributions to Education, No. 204. New York: Teachers College, Columbia University, 1925.

TABLE XXVI

THE GAINS IN MEAN SCORES MADE ON THE U. H. S. TEST BY THE PUPILS IN THE EXPERIMENTAL AND CONTROL GROUPS DURING THE SUMMER VACATION FOLLOWING THE EXPERIMENT

Test	Experimental Group			Control Group		
	Mean Score at end of experiment	Mean Score four months later	Gain*	Mean Score at end of experiment	Mean Score four months later	Gain*
Vocabulary	20.86	21.18	0.32	31.44	29.20	-2.24
Rate	37.69	38.00	0.31	41.52	43.88	2.36
Comprehension . . .	35.64	34.36	-1.28	44.08	42.64	-1.44
Total Score	94.20	93.55	-0.65	117.04	115.72	-1.32

*A minus sign (-) preceding a number in the gain column indicates a loss in score.

vocabulary and comprehension. The differences between average scores on Form 3 and Form 4 were small, but they were somewhat larger than the differences in scores made by the experimental group. The control group had a mean loss of 1.3 points in total score, which loss was greater by 0.7 point than that of the experimental group. The loss, however, was not large enough to be significant, as it could easily be explained by a slight difference in the difficulty of the forms of the test or by the fact that Form 4 was administered to most of the pupils during the first week of the school year before they had become settled into the routine of the school.

The fact that the experimental group did not show a greater loss from June to October than did the control group is highly significant. If the gains which the reading class made during the corrective teaching represented only temporary improvement, the assumption would be that those pupils would lose considerably more, after the reading instruction was discontinued, than would a group that had a slow and presumably normal growth in reading ability at the same time that the experiment was in progress. The findings, however, indicate that the experimental pupils retained their reading skill during the summer vacation as well as did the control pupils.

It should not be concluded from the preceding discussion of mean scores that all of the pupils in the experimental group made approximately the same score in October that they made in June. As a matter of fact, ten pupils lost in total reading score, nine pupils gained, and three pupils neither gained nor lost, but the losses and gains were small in nearly all cases.

The Monroe Standardized Silent Reading Test, Form 2, and the Sangren-Woody Reading Test, Form A, were also given in October to the pupils who had formed the experimental group. In general, the results verified the findings secured with the U. H. S. test.

Summary of Findings

Two experimental seventh-grade classes in corrective reading, taught by one of the instructors in the English department, made significant gains in mean scores on reading tests. These gains were substantially greater than the improvement made by control groups. A ninth-grade class likewise showed definite improvement.

Nearly all the pupils made enough improvement to justify releasing them from further training in reading.

Most of the backward readers derived considerable permanent benefit from the corrective instruction in reading, carried on under regular classroom conditions. Some pupils continued to gain after the teaching was brought to a close, other pupils neither gained nor lost for a time, and some pupils showed a loss, but rarely did a pupil drop back to his reading level before corrective teaching was begun.

CHAPTER VI

THE RELATION BETWEEN RATE OF READING AND SPEED OF ASSOCIATION

The Problem of the Chapter

It will be recalled that the third major purpose of this study was the invention of diagnostic techniques for finding the specific nature of reading deficiencies. In the course of the experimental teaching described in the preceding chapter a need was discovered for intensive study of a number of such problems relating to reading at the junior-high-school level. Some of these were problems of measurement of certain types of reading ability; others dealt with the discovery of reading attitudes and emotional reactions to reading situations; still others were concerned with the relation between reading ability and various other abilities. A thorough investigation of all the problems encountered was, of course, beyond the scope of this study. However, one problem—that of the relationship of rate of reading to speed of association—was investigated in considerable detail. It is used in this study to illustrate the more detailed type of studies essential to the progress of scientific work with retarded readers in the junior high school.

The causes of differences in reading rate.— It has been known for many years that people read silently at unequal rates. During the last two decades there have been many studies of the factors which influence rate of reading. A large number of these factors relate to the conditions under which reading is done. Some of them are (1) the attitude of the reader, (2) the purposes of the reader, (3) the type of reading (silent or oral), (4) the degree of concentration of attention, (5) the nature and the difficulty of the reading material, and (6) the amount of practice in rapid silent reading. Two other factors which appear to influence reading rate in considerable degree are general intelligence and breadth of reading vocabulary.

Perhaps the most thoroughly studied cause of difference among individuals in reading rate is skill in the mechanics of reading. The mature reader has no lip movement in silent reading and vocalizes only to a slight degree. He moves his eyes over the page with quite regular fixations of rather short duration. This type of reader has a great advantage over the one who vocalizes

as he reads and whose eye-movement records show many fixations of long duration, with numerous regressive movements. Some authorities in the field of reading apparently hold that mechanics are practically the only basic factors in rate of silent reading. For example, one of these authorities states in a recent monograph:

"If you are a slow reader your slowness is probably due to one of three reasons:

(1) Your eyes may stop too often in getting across a line of print . . . Pausing too often is the most frequent cause of slow reading.

(2) The next most likely cause for your slowness is that you say the words over to yourself as you read.

(3) Finally, you may have been trained in your elementary and high-school work to 'read every word.' Not every word should be read . . ."¹

The same writer would help slow readers by directing their attention to the mechanics of the process. She has worked out a technique in which she attempts to guide the learner step by step through a series of highly mechanical exercises in eye-movement, with a view to increasing the individual's rate of reading.²

Such emphasis upon the mechanics of reading may be of value for certain individuals, but some cases of slow reading cannot be explained in terms of faulty mechanics nor on the basis of any other cause of differences in reading rate thus far advanced. During the experiment in corrective reading discussed in Chapter V the teacher and the research worker became much interested in finding an explanation for the very slow rate of reading of two of the pupils in the class. These pupils were in the high normal group in intelligence, their knowledge of vocabulary was extensive, they exhibited no lip movement, and their habits of eye movement were quite mature, yet their rates of reading were unquestionably among the slowest in the class. It was also observed that not only did these two pupils read slowly, but they talked slowly and thought slowly as well. This observation raised a question as to whether their tendency to think slowly was basic to their inferior rates of reading. This brought up the whole question of the influence of rate of association of ideas upon speed of reading.

The problem stated.- The problem was to find out if a relationship exists between the speed with which individuals associate

¹Luella C. Pressey, A Manual of Reading Exercises for Freshmen, pp. 3-4. Columbus: the Ohio State University Press, 1928.

²Ibid., pp. 3-23.

ideas with words and their rate of reading. In other words, do some people read rapidly partly because they are able to associate quickly the words of the printed page with ideas, while others read slowly in part because they are essentially slow in their thought processes?

It is evident that the question is of considerable practical significance. Certain standards, based on averages, have been set up for reading rate. All over the country schools are inaugurating programs of remedial reading. One of the purposes of such programs is to bring slow children up to "standard" in rate of reading. The question is raised in this study as to whether effort in that direction may not be futile for certain individuals. Are people destined by native rate of association of ideas to vary considerably in the speed at which they can and do read?

Certain subsidiary steps are associated with the main purpose of the chapter. First, it is necessary to devise a technique for measuring rate of association. Second, if a positive relationship between rate of reading and speed of association is found, it will be desirable to indicate certain practical uses that can be made of a test of rate of association in the diagnosis of rate of reading.

The Technique Used in Studying the Problem

The general plan.- The problem was studied by correlating rate of association with rate of reading scores. Data concerning rate of reading were available in abundance, but it was necessary to gather data on association time.

It would have been highly desirable to have gathered the facts concerning rate of association in a normal reading situation. But this would have been very difficult, if not impossible, because the results would have been influenced by habits of eye movement. It was therefore necessary to devise an artificial testing situation, in which the results were free, as far as possible, from the influence of any factors other than association time.

It should be remembered that any attempt to relate reaction time on a test of this kind to rate of reading involves an important assumption. One assumes that each individual has a general association time which influences alike his rate of making meaningful reactions to continuous written discourse and his speed of reacting to stimuli which are not in context, and that by measuring this general factor as it is manifested in the latter instance, one can secure some data which will help show whether it is influencing rate of reading in the former instance. It seems logical to assume that there is probably no qualitative difference in the association time involved in the two situations.

Criteria which governed the making of the tests.- The following criteria were observed in constructing the association tests.

1. The initial technique may be quite crude. It will be sufficient to indicate general trends. The technique may be refined later, if the preliminary study indicates that the problem is worth pursuing further.

2. The vocabulary used in the tests should be easy, in order that the influence of general intelligence and knowledge of word meaning will be minimized.

3. It is preferable to present the stimuli to the subject visually rather than orally, as this method affords less chance of misunderstanding them. This is particularly true of homonyms.

4. A series of single stimulus words rather than phrases or sentences should be used, as the latter kind of stimulus would involve eye-movements.

5. The number of words included in the tests should be large enough to permit a considerable range in the time required for various individuals to complete the series.

6. In order to develop a technique which is transferable to the average school situation, it is permissible to keep a record of the total time of each individual rather than the time of single reactions, as this can be done with simpler measuring instruments.

7. The test should contain no words which may set up inhibitions because of their peculiar nature.

8. As the best type of test for determining rate of association is not known, it is desirable to experiment with more than one type.

9. Two forms of each test should be devised in order that the reliability of the test can be computed.

10. The technique of measuring rate of association should be refined and improved later on the basis of the experience that is gained in the preliminary experiment.

The types of tests.- A decision was made to use a free-association test and a controlled-association test. These tests are sufficiently different that some of the undesirable features of one type are not duplicated in the other type. For example, speed of reaction to a controlled-association test is probably affected considerably by knowledge of vocabulary, since the subject must know, not only the stimulus word, but also the exact word for which the stimulus calls. On the other hand, rate on a free-association test which is made of very common words seems to be only slightly related to vocabulary. This difference in favor of the free-association test is balanced by the fact that speed in performing it depends, in some degree, upon the ability of the individual to discard his inhibitions and to respond with the first

word that comes to him, regardless of whether or not it sounds well. This feature is avoided in the controlled-association test, as this type of test calls for definite responses.

The free-association test.- It was found that in general the reaction of subjects to words in a free-association test was not very rapid. Therefore, a relatively small number of words gave a wide range in the total time of the various individuals.

Each form of the free-association test consisted of twenty-three words chosen from the three thousand most common words in Thorndike's The Teacher's Word Book. Care was used to select words which would provoke a variety of responses and to avoid those which might set up inhibitions.

The words in the free-association test were shown to the subject one at a time by means of a tachistoscope operated by the examiner. In the administration of the test the subject was seated facing the tachistoscope and the following directions were given:

"Words will be shown to you one at a time through the opening. As soon as you see a word there, say the first thing the word makes you think of. For example, pencil might bring to your mind pen, or paper, or write, or any of several other words. Do the test as quickly as you can."

Other illustrations were given if the subject seemed in doubt concerning what was expected of him.

The examiner then exposed a word. As soon as the subject responded, the examiner closed an electric switch causing the tachistoscope to show the next word. By this technique, the reaction time of the examiner as well as that of the subjects was involved in the results. However, since one individual administered all the tests, it can be assumed that the reaction time of the examiner was practically a constant factor.

The controlled-association test.- The controlled-association test consisted of a test of easy antonyms. Two forms of fifty words each were devised. With three exceptions, the words in the second form consisted of the opposites of those in the first form, but the order of appearance was different.

In the construction of the test, a somewhat longer list of words was prepared and was tried out with a few subjects. The words which caused an unusually long pause before the response were considered too hard and were eliminated.

During the first experiment in which the controlled-association test was used, the words were shown to the subject with a crude type of tachistoscope which required considerable skill in operation. After the preliminary experiment had been carried on a new and improved device for exposing the words was made.

The new tachistoscope consisted of a small, portable wooden

box, in one side of which a small opening was cut. The box was fitted to a detachable base and was so arranged that it could be set up on a table with the slit about on a horizontal line with the subject's eyes. The one hundred words in the test were printed on a strip of motion-picture film. The film was inserted lengthwise through the box in such a position that the words were exposed, one at a time, through the slit. A simple arrangement of cog-wheels and a lever made it possible for the examiner to show the words rapidly and accurately. A sketch of the apparatus is shown in Figure 1.

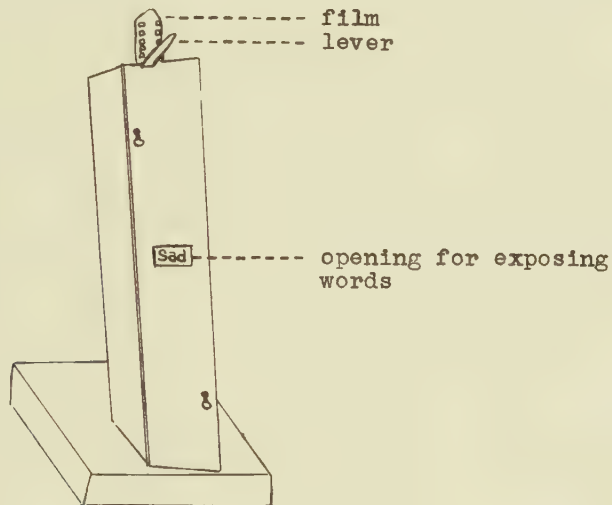


Fig. 1.- The tachistoscope used with the revised controlled association test.

When the test was administered, the subject sat at a table facing the apparatus, which stood on the table at a distance of about eighteen inches from his eyes. The examiner sat at one side in such a position that he could manipulate the apparatus and could observe the subject and read the words which were exposed through the opening. He operated the lever which controlled the movable film with his left hand and held a stop-watch in his right hand.

The directions given to the subject were as follows:

"I have a list of words here and I want you to tell me the opposites of them. I will show a word here (pointing to the slit in the box) and you are to say the opposite as quickly as you can. Then I will show another word and you will say the opposite of that as quickly as possible, and so on through

the list. If you don't think of the opposite word rather quickly just say 'don't remember' and we will go on to the next word."¹

The subjects.- In the first experiments, each of the association tests was administered to fifty seventh-grade pupils of the University High School and to twenty-seven students in the Freshman class of the University of Chicago² during the Winter Quarter, 1931. All of the university students were tested with both tests. Thirty-one of the seventh-grade children took both tests.

The revised test of antonyms was given to sixty-seven ninth-grade pupils of the University High School in the Spring Quarter, 1931, and to forty seventh-grade pupils in the Autumn Quarter, 1931.

The rate of reading scores.- Both groups of students who were used in the preliminary experiments were given rate of reading tests during the Autumn Quarter of 1930. The university freshmen were tested by Mr. I. A. Booker with a reading test of his own construction. The seventh-grade pupils were tested by the writer with five forms of a rate test which he was trying out for the purpose of gathering data concerning the relative difficulty of the forms. Both tests yielded rate scores in terms of number of words read each second. These reading scores were used in the present study.

The ninth-grade pupils who served as subjects in the later experiment were tested for rate of reading with Form 1 of the writer's rate test. The seventh-grade pupils who took the revised test of antonyms were given Form 1 of the writer's rate test and form B of the rate test of the Sangren-Woody Reading Test. As both tests gave a score in terms of number of words covered during continuous reading, each pupil's scores on the two tests were combined. It is probable that the combined score was more reliable than the score on either test alone would have been.

¹The last sentence was necessary because it was found that when it was not included in the directions some children who had difficulty in recalling the opposites of two or three words spent so much time in trying to think of them that their total time was increased greatly. Thus, it was possible for an individual who was really rapid in his associations to make a poor showing in total time because of deficiencies in vocabulary.

²Although the main purpose of this chapter is to devise a diagnostic technique that can be used in the junior high school, this small group of university freshmen has been included in the study in order to have some evidence about the usefulness of the technique with subjects who are relatively mature.

Validity and Reliability of the Association Tests¹

Validity.- The correlations of rate on the controlled-association test with rate on the free-association test has some value as an index of the validity of the tests as measures of association time. Its value in this respect is, of course, limited, as the tests might correlate perfectly even though neither was measuring rate of association. High correlation between the tests would not be conclusive evidence of validity, but low correlation would suggest that one or both of them failed to measure what they were designed to test.

The correlations between the two types of tests were .653 $\pm$.070 for the thirty-one seventh-grade children to whom both tests were given and .678 $\pm$.068 for the twenty-seven university Freshmen. These correlations are not very high for two tests which are used as measures of the same thing, but they are as high as those which are frequently found for tests within a field.

Reliability.- The total time required by each subject to complete both forms of each test was used in the study of the correlation of association time with rate of reading. The reliability of the two forms together, rather than the reliability of each form is, therefore, the more significant fact in this study.

The reliability of the total scores on the forms of each test was estimated by making use of the Spearman-Brown formula for lengthened tests after the reliability of the separate forms had been found. The results are shown in Table XXVII.

Four of the six coefficients of correlation shown in Table XXVII are above .90 and three of them are above .95. These reliability coefficients are surprisingly high. They indicate that those abilities which contributed to the test scores were measured very consistently.²

¹The results of the first experiments with the two types of association tests and of the experiment with the revised test of antonyms will be presented contiguously in this section and the next section, rather than separately, as such an arrangement will lend compactness to the treatment and will facilitate the reading of the data and the interpretations.

²In all cases, Form 2 was administered immediately after Form 1, with an interval of only a few seconds between them. The coefficients of correlation are therefore indicative of the reliability of the technique, only, and cannot be interpreted as suggesting that the abilities measured were relatively fixed and unchangeable. Although the rate at which individuals associate ideas

TABLE XXVII

THE RELIABILITY OF THE 46-WORD FORMS OF THE FREE-ASSOCIATION TEST AND THE 100-WORD FORMS OF THE TEST OF ANTONYMS

Test	Seventh Grade Pupils	Ninth Grade Pupils	University Freshmen
Free-association test . . .	.957 $\pm$.008	---	.895 $\pm$.026
Antonyms (original test) .	.949 $\pm$.011	---	.963 $\pm$.010
Antonyms (revised test) . .	.973 $\pm$.006	.887 $\pm$.018	---

The average rate of performing the tests.- The mean scores made by the four groups of students on the association tests are shown in Table XXVIII.

TABLE XXVIII

THE MEAN NUMBER OF SECONDS REQUIRED TO PERFORM THE ASSOCIATION TESTS

Group	Test of Antonyms			Free-Association Test		
	Form 1	Form 2	Both Forms	Form 1	Form 2	Both Forms
First Experiment						
Seventh-grade pupils	136.2	137.9	274.1	89.0	83.4	172.4
University Freshmen	95.6	91.2	186.8	68.9	64.0	132.9
Second Experiment						
Seventh-grade pupils	121.5	123.9	245.4	--	--	--
Ninth-grade pupils	103.5	102.2	205.7	--	--	--

In the first experiment, the university Freshmen were considerably more rapid than the seventh-grade pupils, especially on the test of opposites. In the second experiment, the ninth-grade pupils were faster than the seventh-grade pupils, although not as

with words is probably relatively stable, it is quite possible that if the forms of the test had been administered at different sittings the correlations between them would have been lower.

rapid as were the university Freshmen in the first experiment.¹ Whatever was measured by these tests evidently developed with maturity. This finding may raise a question as to whether these tests really measured association time, as speed of reaction probably makes no marked increase in the few years that separate the ages of the two groups. However, it will be shown later in this chapter that rate on these tests was correlated with both intelligence and vocabulary, which is perhaps a sufficient reason for the difference in favor of the more mature group.

The free-association test showed somewhat more practice effect than did the controlled-association test. Although Form 1 of the test of antonyms was administered first in all cases, the average time of both groups of seventh-grade pupils was slightly longer on Form 2 than on Form 1. Practice did not materially influence the results of either test except in the case of a few individuals.

Correlation of Reading Rate with Speed of Association

Association time correlated with rate of reading.- The rate of performance on the free-association test, the first opposites test, and the revised test of opposites was correlated with rate of reading for the students in each of the groups who took the tests. The coefficients of correlation are shown in Table XXIX.

All six correlations between the association tests and the rate of reading tests are high enough to be significant. They suggest that association time is an important factor in rate of reading.

It is possible, however, that the rates on these tests are correlated, not because both types of test are influenced by association time, but because of the effect of some other ability or abilities. For example, these tests may be correlated highly with intelligence and thus may gain their apparent correlation from this relationship. The question of the correlation of the association tests and the rate of reading tests with intelligence will now be investigated.

¹The difference between the mean rate of the seventh-grade pupils in the first experiment and the mean rate of the seventh-grade pupils in the second experiment was probably due to the difference in the apparatus for exposing the words. The tachistoscope used in the second experiment worked much smoother and was considerably less distracting to the subjects than the one used in the first experiment.

TABLE XXIX

CORRELATION OF SPEED OF ASSOCIATION WITH RATE OF READING

Test	Reading Rate Seventh Grade Pupils	Reading Rate Ninth Grade Pupils	Reading Rate University Freshmen
Free-Association Test . .	.528 $\pm$.069	---	.487 $\pm$.10
Antonyms (original test)	.527 $\pm$.069	---	.565 $\pm$.091
Antonyms (revised test) .	.558 $\pm$.074	.392 $\pm$.071	---

Relation of association time and rate of reading to intelligence.- Each group of students took an intelligence test at the time of their entrance to the University. The Otis Self-Administering Test of Mental Ability, Higher Examination, was given to the high-school pupils and the Thurstone Psychological Examination was administered to the university Freshmen. The rate of performance on the association tests and the rate of reading scores of the students were correlated with their point score on the Otis test or their percentile rank on the Thurstone test. The correlations are presented in Table XXX.¹

All of the correlations in Table XXX except the one between the rate of reading and the intelligence scores of the ninth-grade pupils appear to be significant, and several of the coefficients are quite high. It should not be concluded, however, that the correlations indicate the extent to which rate of association and rate of reading are influenced by intelligence. The mental tests used in these correlations were timed tests, and the scores made by the pupils who took them were doubtless dependent, in part, upon their reaction time and their rate of reading. In other words, the intelligence of these students, if it is defined as the scores they made on a mental test within a given time limit, was to some degree a function of rate of association and reading rate. It may be assumed, therefore, that the real influence which mental ability had upon rate scores in the association tests and the reading tests

¹It was necessary in the first experiment to compute two correlations for rate of reading and the Otis test, as the group of seventh-grade pupils to whom the test of opposites was given was not identical with the group to whom the free-association test was administered.

TABLE XXX

CORRELATION OF ASSOCIATION TIME AND RATE OF READING
WITH INTELLIGENCE

Tests	Otis Point Score			Thurstone Percentile Rank
	7th Gr. Pup. who took opp. test	7th Gr. Pup. who took fr-assoc t.	9th Grade Pupils	University Freshmen
First Experiment Opposites (original)	.371 $\pm$.082	---	---	.688 $\pm$.072
Free-association	---	.488 $\pm$.072	---	.412 $\pm$.112
Rate of Reading	.524 $\pm$.069	.434 $\pm$.077	---	.590 $\pm$.089
Second Experiment Opposites (revised)	.656 $\pm$.061	---	.510 $\pm$.062	---
Rate of Reading	.639 $\pm$.063	---	.197 $\pm$.080	---

was somewhat less than the correlation coefficients indicate.

Correlation of speed of association and rate of reading when intelligence is held constant.- Thus far in this chapter it has been found that rate of association appears to be correlated with rate of reading and that both factors seem to be related to intelligence. These findings point to a conclusion that the relationship between association time and reading rate is due, in part, to the influence of intelligence. This raises a question as to whether or not there would be a significant correlation between association time and rate of reading if intelligence were not a factor in the relationship. An answer to this question might be found readily by the method of partial correlation if the correlations shown in Table XXX represented only the effect of intelligence upon the rate scores, rather than an interaction of influences. But, as has already been pointed out, the paths of influence go in both directions. Therefore, if rate of reading is correlated with rate of association, while intelligence is held constant, too much will be taken out and the resulting coefficients will be lower than the true relationship.

A certain value, however, will be derived from computing a partial correlation of this kind, if it is kept in mind that too

much is taken out by this method. If the coefficient is so low as to be insignificant, this will not prove that there is no direct relationship between rate of association and rate of reading. If, however, there is significant correlation despite the fact that too much has been taken out, this will be strong evidence that speed of association is connected with reading rate directly and not by virtue of the relation which both factors bear to intelligence. The correlation of rate on the three association tests with rate of reading, with intelligence held constant, is shown in Table XXXI.

TABLE XXXI

PARTIAL CORRELATION OF SPEED OF ASSOCIATION WITH RATE OF
READING WHEN INTELLIGENCE IS HELD CONSTANT

Test	Reading Rate Seventh Grade Pupils	Reading Rate Ninth Grade Pupils	Reading Rate University Freshmen
Free-Association test . .	.403 $\pm$.080	---	.331 $\pm$.120
Antonyms (original test)	.421 $\pm$.078	---	.271 $\pm$.122
Antonyms (revised test) .	.239 $\pm$.100	.346 $\pm$.073	---

All of the correlation coefficients were considerably reduced by partialling out intelligence. The amount of reduction for the seventh-grade pupils was .125, .106, and .319; for the ninth-grade pupils, .046; for the university Freshmen, .156 and .294.

The partial correlations of the free-association test and the original test of antonyms with rate of reading are significant for the seventh-grade pupils, since they are more than four times their probable error. For the twenty-seven university Freshmen, the partial correlation of rate of reading with speed on the test of antonyms is only 2.2 times as large as the probable error; the corresponding correlation of rate of reading with speed of free association is about 2.8 times as large as its probable error. These correlations may be regarded as somewhat significant, although the evidence of relationship which they offer is slight. The correlation of speed on the revised test of antonyms and reading rate scores of the seventh-grade pupils is scarcely significant, as it is only 2.4 times its probable error. The same kind of correlation, using the scores of the ninth-grade pupils, is more

than four times its probable error and is therefore significant. Thus, three of the coefficients of correlation are significant, while the other three are of doubtful value. In general, when all six of the partial coefficients are considered, it appears that speed of association has some connection with rate of reading which is independent of mental ability.

Correlation of association time and rate of reading with vocabulary.- As was pointed out in the first section of this chapter, the knowledge of word-meaning possessed by pupils has a bearing on the rate at which they read. It has also been found that the quality of responses secured on an association test is closely related to extent of vocabulary.¹ If rate of performing an association test is connected to a like degree with vocabulary, then it may be that understanding of word-meaning rather than association time is responsible for the correlations between the association tests and the tests of reading rate.

The question was investigated by correlating the speed on the association tests and the rate of reading scores with the scores made by the seventh-grade children on the Inglis Test of English Vocabulary, which the pupils took when they entered the University High School. The Inglis test contains 150 words and is not timed. The coefficients of correlation are shown in Table XXXII.²

¹In a recent study, the writer found that the correlation between the comprehending responses to twenty words in a free-association test and scores on the Inglis Test of English Vocabulary for 81 seventh-grade pupils was $.677 \pm .046$.

²It was impracticable to use the scores of the college Freshmen in this part of the study because of the nature of the vocabulary test which they had taken. This vocabulary test used was Part I of the Minnesota Reading Examination for College Students. It is a multiple-choice test, containing 100 words, and is standardized with a working time of six minutes. Only the very exceptional individual can complete it within the time allowed. Both rate of reading and reaction time are undoubtedly important factors in the score which an individual makes on this test. If one should utilize these results in the partial correlation of reading rate and association time, he would tend to hold constant the factors whose relationship he wished to investigate.

Neither was it feasible to make use of the scores of the ninth-grade pupils on the Inglis Test of English Vocabulary. A part of the pupils took the test when they entered the seventh grade, while the rest of the group took it when they entered the ninth grade, and the scores of the two groups were not comparable.

TABLE XXXII

CORRELATION OF ASSOCIATION TIME AND RATE OF READING
WITH VOCABULARY

Tests	Ingllis Vocabulary Test	
	7th Gr. Pupils who took opp. test	7th Gr. Pupils who took free- assoc. test.
First Experiment		
Opposites Test (original)	.366 $\pm$.080	---
Free-Association Test	---	.245 $\pm$.089
Rate of Reading	.510 $\pm$.071	.474 $\pm$.074
Second Experiment		
Opposites Test (revised)	.513 $\pm$.078	---
Rate of Reading	.633 $\pm$.064	---

In general, the foregoing coefficients of correlation are quite similar in size to those which were found when intelligence scores were correlated with rate of reading and speed of association. Five of the six correlations are significant; the sixth, which is the correlation between vocabulary scores and speed on the free-association test may be significant.

Correlation of speed of association and rate of reading when vocabulary is held constant.- As all pupils were allowed to complete the vocabulary test, presumably the scores were not affected by the rate at which various individuals worked. Therefore, a correlation between rate of association and reading rate, with vocabulary held constant by partial correlation, should approximate the true relationship which exists between these abilities among individuals who possess equivalent knowledge of word-meaning. The partial correlation of each of the association tests with rate of reading, when vocabulary is held constant, is shown in Table XXXIII.

The amounts of reduction in the correlations between the association tests and rate of reading that resulted from partialling out vocabulary were .103, .045, and .207, respectively. Notwithstanding these reductions, the first coefficient of correlation is more than five times its probable error, the second one is nearly seven times its probable error, and the third one is almost four times its probable error. The correlations indicate that there is a relation between rate of reading and association time

TABLE XXXIII

PARTIAL CORRELATION OF SPEED OF ASSOCIATION WITH RATE OF
READING WHEN VOCABULARY IS HELD CONSTANT

Tests Correlated	<u>r</u>	P.E.
Original test of opposites and rate of reading . . .	.425 <u>±</u>	.078
Free-association test and rate of reading	.482 <u>±</u>	.072
Revised test of opposites and rate of reading . . .	.351 <u>±</u>	.093

which is independent of vocabulary.

Summary of Facts about Association

Time and Rate of Reading

Clear evidence of a direct relation between rate of reading and speed of association was found with four groups of pupils. The effect of eye-movement was eliminated by the conditions of the study, and the influence of mental ability and that of knowledge of vocabulary were each taken out by partial correlation; nevertheless, the evidence of relationship persisted. This relationship, while it was not marked, was positive and was high enough to be significant.

How the Technique Developed in this

Chapter Can Be Used

This procedure for studying the rate at which individuals associate ideas with words should be of considerable practical value to teachers of reading in the junior high school. The method of measuring association time is so simple that a teacher without special technical training can use it. A tachistoscope similar to the one used with the revised test of antonyms can be made at a small cost and can be operated by anyone with a little practice.

If a test of rate of association is used in conjunction with a reliable battery of intelligence, reading, and vocabulary tests, it is possible to make a rather thorough preliminary diagnosis of a pupil's reading deficiencies and to prescribe a program of corrective and remedial treatment. The way in which this can be done will now be illustrated with an abbreviated analysis of the reading status of three pupils in the seventh grade of the University High School.

Pupil One is in the lower quartile of his class in rate of reading. His I.Q. is 108. He is above the median of his class in

comprehension of reading material. His mastery of mechanics is good. He is above the median of his grade in vocabulary. He has an excellent attitude toward the work of the school and gives almost perfect attention in class. He was one of the slowest of the seventh-grade pupils in his reactions to the test of antonyms. It is probable that his long reaction time is one of the most important factors in his slow rate of reading. Probably he should not be urged to try to make any great gain in reading rate.¹ Since he comprehends what he reads very well, he can undoubtedly do the work of the school satisfactorily, even though he reads slowly.

Pupil Two is considerably above the median of her class in rate scores on reading tests, but her comprehension scores are quite low. Her I.Q. is 105. She is in the lower quartile of her class in vocabulary. Her attitude and attention are satisfactory. She is in the lower quartile of the group in rate on the test of antonyms. Apparently, she has neither the knowledge of word-meaning nor the speed of association necessary to be a rapid reader. It is probable that her present rate of moving her eyes over the page is too rapid for her slow rate of association and that this decreases comprehension. She should be advised to read more slowly.

Pupil Three is one of the slowest readers in the seventh grade. His I.Q. is 95. He is below the median of the grade in power of comprehension and in vocabulary. There is some evidence of lip movement and vocalization. His rate of association is slightly above the mean. It is probable that his exceedingly slow rate is not dependent upon native rate of reaction but rather upon faulty mastery of mechanics and limited knowledge of the meaning of words. It is obvious that the first steps in treatment are to attempt to correct the lack of mastery of mechanics and to help the pupil to acquire a larger stock of sight words. Once progress has been made along those lines, the pupil should be required to read under timed conditions and should be encouraged to "speed up" his reading, as his rate of performing the association test indicates that he can learn to read considerably more rapidly than he now does.

¹The assumption is made in these analyses that rate of association cannot be increased through specific practice. It is possible that this assumption should be modified. This problem needs further investigation.

CHAPTER VII

CONCLUSIONS AND INTERPRETATION

This study has been concerned with the measurement and improvement of reading at the junior-high-school level. Three fundamental problems have been attacked: (1) the development of improved means of measuring silent reading achievement; (2) the development of corrective measures that are economical of time and that can be used by classroom teachers; and (3) the development of special diagnostic techniques that will supplement those now used in determining the specific nature of reading deficiencies. The constructive work on these problems has been presented in the foregoing chapters. The purpose of this chapter is to point out the general conclusions that can be drawn from the specific findings, to indicate the contributions of this study to the problem of the teaching of reading at the junior-high-school level, and to suggest certain lines of investigation along which future studies should be directed.

The Measurement of Silent Reading

A number of excellent reading tests had been developed before this study was begun. However, no one test was satisfactory for use in determining the achievements and needs of groups of junior high-school pupils and in corrective reading programs of several months' duration. A test battery was needed that would have the following characteristics: (1) it should measure several aspects of reading ability in order to be of diagnostic value; (2) it should consist of several forms so that measurements of reading achievement could be made at intervals during the year; and (3) it should be sufficiently valid and reliable to form a basis for conclusions concerning pupil progress in reading. This need could be partially filled by employing a battery of independent standardized tests, but the time used in administering and scoring them seriously limited the practicability of utilizing such a battery of tests.

Accordingly, the task of devising and partially standardizing tests of rate of reading, reading vocabulary, and comprehension, in four forms, was undertaken in the first part of this study. The tests, in their present form, represent the outcome of a process of selection of material, trial, and revision, in which

the scores of approximately 3,000 pupils were utilized. The following conclusions concerning the tests were reached as a result of detailed findings based upon the experimental use of the tests.

Conclusions about the reading tests.-

1. The tests constructed in this study are valid instruments for measuring the reading ability of junior-high-school pupils. More specifically, it appears that:

(a) Each of the tests shows rather high validity when it is correlated with a criterion of the particular kind of reading ability measured by the test.

(b) The battery as a whole has high validity for the purpose of discovering inferior readers.

(c) The tests have considerable validity for the prognosis of success in school.

(d) All of the individual items in the tests show positive validity according to the method of validating questions used in this study.

2. In general, the tests are quite reliable, as compared with the reliability of other achievement tests. Correlations between different forms of the tests indicate that the reliability coefficients are approximately as follows:

Rate	.80+
Vocabulary	.85
Comprehension, Part I	.65
Comprehension, Part II	.73
Comprehension, Total	.80
Composite Score	.92

The reliability of the composite score appears to be as high as that of any standardized reading test used in the junior high school.

3. The various forms of the battery of new tests are practically equivalent. Small differences in mean scores and in distributions of the scores were found, but those differences were, in general, insignificant.

4. The battery of tests meets satisfactorily the other commonly accepted criteria of a good test, such as objectivity, discrimination between reading ability in successive grades, and ease and rapidity of administration and scoring.

Pedagogical implications.- The following implications resulting from the study of the measurement of reading rate, vocabulary, and comprehension have a relationship to teaching procedure:

1. A battery of tests in which independent measures of rate of reading, reading vocabulary, and reading comprehension are obtained is sufficiently diagnostic to furnish a starting point

for setting up a program of corrective reading.

2. It is possible to measure reliably these three aspects of reading achievement within the time limit of one class period of fifty minutes.

3. A well-rounded program of testing in reading, extending over a period of a year, may be carried on with a battery of reading tests containing four forms.

Problems of measurement that need further study.- Tests for ability to do specific kinds of reading should be developed much further than they are at present. For instance, the means of testing the ability of junior high-school pupils to follow directions contained in reading material are highly inadequate. There are a few standardized tests, such as the Burgess Scale for Measuring Ability in Silent Reading and the Following-Directions part of the Sangren-Woody Reading Test, but they are quite elementary and the situations which they set up are artificial. There are no tests for finding out how well a pupil can read and follow the kind of directions he meets in a variety of life situations, such as directions for setting up apparatus in a science experiment, for filling out reports, for assembling the parts of a machine, for a program of the day's work, for an English assignment, et cetera. The teacher can devise informal tests for following directions, but these tests do not tell one how the ability of the pupils tested compares with that of pupils in general, nor do they furnish a reliable basis for measuring progress in this skill. A long and carefully designed test for following directions, containing several forms and duplicating actual situations such as pupils frequently meet in their work would be a noteworthy contribution to the field of corrective reading.

Similarly, valid and reliable tests for ability to follow the writer's organization and to find the main points of reading selections are needed. Few abilities are more closely related to successful study than ability to understand the organization of the material read and to distinguish between main points and sub-points. Probably no phase of a corrective-reading program needs more emphasis than the training of the pupils to do these types of reading. Nevertheless, since there are no satisfactory standardized tests for these kinds of reading, training cannot be directed with maximum effectiveness toward the points of weakness because of the fact that the teacher must depend almost entirely upon informal testing material. The need in these fields is for rather long tests with generous time limits. The tests are needed for diagnostic purposes. The error of sampling in a short test is so large in comparison with the scores involved that individual

diagnosis based on the results is of little value.

Corrective Teaching of Reading

Preceding this study much of the experimental work in teaching retarded readers at the junior-high-school level had been individual in nature. Consequently a majority of the techniques for remedial teaching of reading that had been described were suitable for use with an individual rather than with a group. Several experiments in teaching groups of inferior readers had been carried on at various levels of the school with encouraging results, but most of these studies were of short duration, a number of them did not utilize an adequate testing program, and nearly all of them neglected entirely the question of permanence of gains made during corrective teaching.

An investigation was therefore undertaken in the University High School with the following major purposes in view: (1) to study the feasibility of teaching corrective reading to groups of pupils at the junior high-school level; (2) to develop procedures to be used in group corrective teaching; and (3) to investigate the permanence of the gains made during the corrective training. The following conclusions are justified by the results of the experiment.

Conclusions about corrective teaching of reading.-

1. The reading ability of three groups¹ of pupils at the junior-high-school level, who were normal in intelligence but retarded in reading, was improved through training carried on by teachers of English under normal classroom conditions. Each of the groups made decidedly larger gains in reading than did control groups and also showed much more than the normal amount of improvement on standardized tests.

2. In this experiment three kinds of procedure for corrective teaching of reading were tried with salutary results in each case. One group was given intensive training on certain specific skills during a period of six weeks; in another group the training period lasted one semester and the study of reading was kept subsidiary to mastery of subject matter; with the third group the work in reading was organized into three periods of intensive training on specific skills. The three training periods in the last experiment were distributed throughout a school year. No reliable conclusion concerning the relative value of these procedures can be drawn from this study; improvement among the majority of pupils

¹The work in one of the groups was not an integral part of this study, as the writer did not participate in it.

was secured with each of them.

3. The evidence secured with tests administered several months after the formal teaching of reading was discontinued indicated that the reading gains made during the corrective teaching were relatively permanent ones. Mean scores on the tests administered after an interval of several months were not significantly lower than the mean scores on corresponding forms of the same tests administered at the end of the training period. Some pupils gained a few points, while others lost in score, but very few pupils returned to the reading level that they were on when corrective teaching was begun.

Pedagogical implications.- The conclusions reached following the experimental teaching of reading have three important implications for reading teachers at the high-school level.

1. The reading ability of the majority of retarded readers in the junior high school can be improved through corrective training in reading under regular classroom conditions. The teacher need not be a specialist in reading.

2. The specific kinds of training used in this experiment are satisfactory for a group of inferior readers of normal mental ability. It should not be inferred, however, that they will work with any group.

3. Gains made during the corrective teaching of reading at the junior-high-school level are in most cases retained by the pupils, according to the results of reading tests given several months after training has been discontinued.

Problems for future study.- During the experiment in corrective reading a number of important unsolved problems were encountered. Among those which especially merit investigation are the following:

1. The question of how to discover quickly and accurately the reading attitudes of children and the reasons back of those attitudes is oftentimes baffling to one who is doing corrective teaching. Occasionally, retarded readers are encountered whose only real difficulty is that they neither want to read nor like to read. They are either lacking in the intellectual curiosity that usually motivates pupils to read or they prefer to get their experience through concrete and direct channels. These children are among the most difficult reading cases, because they gradually drop behind the other children from lack of reading practice and yet there is no definite starting point for corrective teaching. Various procedures must be tried with each individual until some effective means of stimulating the pupil's interest in reading is found. There is unquestionably much need for improvement in techniques of studying the interests and attitudes of individual

pupils. Although this is not a reading problem, primarily, yet it is vitally related to the success of the reading teacher in doing certain types of corrective work.

2. A problem that especially needs the attention of research workers is that of collecting or developing suitable reading materials for use in remedial and corrective work at the junior-high-school level. The reading materials that are sufficiently mature in content to absorb the attention of pupils in the junior high school often make use of a vocabulary that is, in part, unintelligible to inferior readers. On the other hand, materials that are free from vocabulary difficulties are usually too juvenile in theme to interest pupils of junior-high-school age. It may be that the solution of the problem will necessitate the rewriting of a considerable portion of the reading materials especially favored by junior-high-school pupils in simpler language.

3. When pupils are retested several months after they have had corrective training in reading, it is found that they fall into three groups: (1) pupils who have continued to improve in reading after the end of the period of instruction; (2) pupils who have neither gained nor lost in reading skill; (3) pupils who are less capable readers than they were when teaching was discontinued. What are the factors in the reading handicaps of the pupils and in the training that is given to them which make for lasting improvement in the majority of cases, but for only temporary gain followed by ultimate loss in other cases, and what steps are needed in order to keep the latter pupils from regressing toward their former inferior habits of reading?

4. The present study has shown that two widely different procedures are effective with retarded readers. Gains on reading tests were secured as a result of intensive training on reading skills; gains were likewise made when the training in reading was held secondary to the study of the regular content of the course. Since the major emphasis in the study was upon intensive training in specific reading skills and as the tests used in the different experimental groups were not identical, conclusions about which procedure was the more effective cannot be drawn from this study. The question of the relative value of intensive training in reading regardless of organization of subject matter and of correlation of reading with the study of content subjects is a particularly appropriate field for further investigation.

Relation of Rate of Reading to Speed of Association

Scientific techniques for finding the specific nature of reading difficulties constitute a comparatively new field. Practi-

cally all these techniques have originated within the last twenty years. The development of the apparatus for photographing eye-movements has given great impetus to studies in this field. Some of the more recently constructed tests have likewise afforded means by which the diagnosis of reading deficiencies can be made. Many additional techniques of this kind are needed.

Although the more important purpose of the present investigation was not concerned directly with the development of techniques for analyzing reading deficiencies, one problem of this kind was studied in connection with the experiment in corrective reading and was presented in this thesis to illustrate a type of detailed analytical study that is needed. The problem chosen for investigation was the relation between rate of reading and speed of association.

The problem was attacked by devising tests of speed of association and by administering these tests together with tests of rate of reading to various groups of pupils and correlating the scores. This procedure resulted in the following conclusions:

Conclusions about the relation of rate of reading to speed of association.-

1. There is clear evidence of a significant correlation between speed of association and rate of reading. It was found that four of six correlations between rate on association tests and rate on reading tests were above .50.

2. When intelligence and vocabulary were each held constant by the method of partial correlation, the relationship between speed of association and rate of reading persisted. This relationship was not marked, but it was positive and was high enough to be significant.

Pedagogical implications.- Speed of association bears a sufficiently close relation to rate of reading that the rate with which pupils associate ideas should be considered by the teacher of reading in diagnosing abnormal rates of reading.

Problems for future investigation.- The study of the relation between rate of reading and speed of association reported in this thesis should be regarded as one which makes a preliminary attack upon the problem. Two important questions growing out of the study are the following:

1. Although this study shows that rate of reading is related to speed of association, the study does not show whether rate of making the kind of associations used in reading can be increased by specific practice. This is a fundamental problem as it is basic to prescription of treatment for pupils who are slow in making associations. In this study, the tentative position has been taken that, in the absence of evidence that speed of association

can be increased by training, it is not advisable to place pressure upon slow readers who are also slow in associating ideas to get them to read more rapidly. Additional investigation may show that this position is not tenable.

2. The procedure for studying rate of association used in this study is somewhat limited in value when it is used with pupils who have very meager vocabularies. It is possible, too, that it is not valid for use with pupils in grades below the junior-high-school level. What is a valid procedure for studying the rate of association of pupils in the lower grades, or of pupils in the upper grades who have very small vocabularies?

The Need for Continuous Investigation

It has been shown in this chapter that our knowledge of reading problems at the junior-high-school level has been extended by this investigation. One of the major purposes of the study would be defeated, however, if any impression were left with the reader that the type of work carried on in the study is now complete. Rather, the point merits emphasis that this study is but a part of a continuous investigation of problems relating to reading at the high-school level that must be conducted over a period of many years. It is to be hoped that the specific findings made during the investigation and the unsolved problems indicated in the foregoing pages will help to stimulate others to search for practicable procedures and plans for analyzing reading deficiencies and for improving the reading habits of backward pupils.

B I B L I O G R A P H Y

BIBLIOGRAPHY

Articles

- Alderman, Grover H. "Improving Comprehension Ability in Silent Reading," Journal of Educational Research, XIII (January, 1926), 11-21.
- Alexander, Fred M. "Reading in Newport News High School," Journal of Education, CXIII (March 2, 1931), 248-51.
- Beeby, D. J. "Diagnostic and Remedial Work in Silent Reading," Chicago Schools Journal, VIII (February, 1926), 204-12.
- Bruene, Elizabeth. "Effect of the Summer Vacation on the Achievement of Pupils in the Fourth, Fifth and Sixth Grades," Journal of Educational Research, XVIII (November, 1928), 309-14.
- Burks, Jesse D., and Stone, Clarence R. "Relative Effectiveness of Two Different Plans of Training in Silent Reading," Elementary School Journal, XXIX (February, 1929), 431-36.
- Carney, Elizabeth, and Leuenberger, Claire S. "Experiment in Remedial Reading in Senior High School," High School Teacher, VII (February, 1931), 71-72.
- Carter, Homer L. J. "Disabilities in Reading," Elementary School Journal, XXXI (October, 1930), 120-31.
- Current, W. F., and Ruch, G. M. "Further Studies of the Reliability of Reading Tests," Journal of Educational Psychology, XV (1926), 476-81.
- Davis, Georgia. "Procedures Effective in Improving Pupils of Poor Reading Ability in Regular Reading Classes," Elementary School Journal, XXXI (January, 1931), 336-48.
- Drainfield, Edgar J. "A Technique for Teaching Silent Reading," Teachers College Record, XXVI (May, 1925), 740-52.
- English Department of the University High School. "Differentiating Instruction in Ninth Grade English," School Review, XXVII (December, 1919), 772-88.
- Eurich, Alvin C. "The Relation of Speed of Reading to Comprehension," School and Society, XXXII (September 20, 1930), 404-6.
- Garrison, S. C., and Robertson, M. S. "Reliability of the Thorndike-McCall Reading Scale," Peabody Journal of Education, IV (1926), 162-64.
- Gates, A. I. "An Experimental Study of Reading and Reading Tests," Journal of Educational Psychology, XII (1921), 445.

- Gray, William S. "Case Studies of Reading Deficiencies in Junior High Schools," Journal of Educational Research, X (1924), 132-40.
- Gray, William S. "Summary of Reading Investigations (July 1, 1929 to June 30, 1930)," Elementary School Journal, XXXI (March and April, 1931), 531-46, 592-606.
- Gray, William S. "Summary of Reading Investigations (July 1, 1930 to June 30, 1931)," Elementary School Journal, XXXII (February, March, and April, 1932), 447-463, 510-520, 587-594.
- Hilliard, George H., and Barnes, Marcillene. "The Effect of Specific Drill on Reading Ability," Elementary School Journal, XXXI (February, 1931), 417-26.
- Hogoboom, Grace W. "Remedial Reading for the Slow-Learning Pupils of the Junior High School," Los Angeles Educational Research Bulletin, X (April, 1931), 13-16.
- Jacobson, P. B., and Vandusen, E. C. "Remedial Instruction in Reading in the Ninth Grade," School Review, XXXVIII (February, 1930), 142-6.
- Keener, E. E. "How to Study in High School," Chicago Schools Journal, XI (January, 1929), 167-69.
- Leonhardy, Alma; Hogoboom, Grace W.; Henry, Mary Bess; Brown, Rubetta D.; and Farmer, Elinor R. "What is Done to Improve the Reading Ability of Pupils in Los Angeles Junior High Schools," Los Angeles Educational Research Bulletin, VII (October-November, 1927), 4-7, 10-11.
- Lowry, Ellsworth. "A Silent Reading Experiment in Rural and Village Schools," Elementary School Journal, XXVI (September, 1925), 42-49.
- McCallister, James M. "The Effectiveness of Remedial Instruction in Reading in the Junior High School," School Review, XXXIX (February, 1931), 97-111.
- McCallister, James M. "Guiding Pupils' Reading Activities in the Study of the Content Subjects," Elementary School Journal, XXXI (December, 1930), 271-84.
- McCallister, James M. "Reading Difficulties in Studying Content Subjects," Elementary School Journal, XXXI (November, 1930), 191-201.
- McCallister, James M. "Character and Causes of Retardation in Reading among Pupils of the Seventh and Eighth Grades," Elementary School Journal, XXXI (September, 1930), 35-43.
- McCarty, Pearl Sands. "Increasing Comprehension in Silent Reading," School Review, XXXIX (December, 1931), 758-66.
- Middlebrook, Pearl H. "Remedial Work in Reading," Elementary School Journal, XXVII (December, 1926), 305-6.
- Miles, Dudley H. "Can the High School Pupil Improve His Reading Ability?" Journal of Educational Research, XIV (September, 1926), 88-98.

- Morrison, J. C. "What Effect Has the Summer Vacation on Children's Learning Ability and Ability to Learn?" Educational Research Bulletin, III (October 1, 1924), 245-49. Columbus, Ohio: Bureau of Educational Research, Ohio State University.
- Mosher, R. M. "A Further Note on the Reliability of Reading Tests," Journal of Educational Psychology, XIX (1928), 272-74.
- Nelson, M. J. "The Abilities and Achievements of Elementary School Pupils before and after a Vacation," School and Society, XXVIII (September 22, 1928), 371-72.
- O'Brien, Ida. "A Comparison of the Use of Intensive Training and of Wide Reading in the Improvement of Reading," Educational Method, X (March, 1931), 346-49.
- Otto, Henry J. "Remedial Instruction in Reading," Elementary School Journal, XXVIII (January, 1928), 353-61.
- Rhoades, Luke C. "Systematic Remedial Training for Pupils of Superior Reading Ability," Elementary School Journal, XXIX (June, 1929), 771-73.
- Rhodes, E. N. "Technique in Teaching Silent Reading," Elementary School Journal, XXIII (December, 1922), 296-302.
- Sangren, Paul V. "Methods of Diagnosis in Reading," Elementary English Review, VII (April, 1930), 105.
- Shank, Spencer. "Student Responses in the Measurement of Reading Comprehension," Journal of Educational Research, XXII (September, 1930), 119-29.
- Shepherd, Edith E. "Some Silent Reading Lessons in Junior High School English," School Review, XXIX (March, 1921), 206-15.
- Shields, James M. "Teaching Reading Through Ability Grouping," Journal of Educational Method, VII (September-October, 1927), 7-10.
- Simpson, I. Jewell, and Stern, Bessie C. "Improving Instruction in Reading," Elementary School Journal, XXV (April, 1925), 594-606.
- Springstead, Clara B. "Improving Speed and Comprehension in Reading," Journal of Educational Method, V (October, 1925), 48-52.
- Sultzer, M. F. "Remedial Measures," National Education Association Journal, XIX (February, 1930), 43-4.
- Sutherland, A. H. "Correcting School Disabilities in Reading," Elementary School Journal, XXIII (September, 1922), 37-42.
- Thorndike, E. L., and Symonds, Percival. "Difficulty, Reliability, and Grade Achievements in a Test of English Vocabulary," Teachers College Record, XXIV (November, 1923), 438-45.
- Vandusen, E. C. "Remedial Instruction in Reading in the Ninth Grade," School Review, XXXVIII (February, 1930), 142-46.

Yoakum, Gerald A., and Truby, Charlotte. "The Effects of Certain Practice Exercises in Reading," University of Pittsburgh School of Education Journal, I (January-February, 1926), 60-61.

Books

Hill, Howard C. Community Civics. Boston: Ginn and Company, 1928.

Hill, Howard C. Readings in Community Life. Boston: Ginn and Company, 1930.

Holzinger, Karl J. Statistical Methods for Students of Education. Boston: Ginn and Company, 1928.

Holzinger, Karl J., and Mitchell, Blythe C. Exercise Manual in Statistics. Boston: Ginn and Company, 1929.

Lyman, Rollo L., and Hill, Howard C. Literature and Living. New York: Charles Scribner's Sons, 1925.

McCall, William A., and Crabbs, Lelah Mae. Standard Test Lessons in Reading, Book V. New York: Bureau of Publications, Teachers College, Columbia University, 1926.

Odell, C. W. Traditional Examinations and New Type Tests. New York: The Century Company, 1928.

Pitkin, Walter B. The Art of Rapid Silent Reading, pp. 64-73. New York: McGraw-Hill Book Company, 1929.

Pressey, Luella C. A Manual of Reading Exercises for Freshmen. Columbus: The Ohio State University Press, 1928.

Romanes, G. J. Mental Evolution in Animals, pp. 136-37. New York: D. Appleton and Company, 1884.

Ruch, G. M. The Objective or New Type Examination. Chicago: Scott, Foresman and Company, 1929.

Sangren, Paul V. The Measurement of Achievement in Silent Reading. Kalamazoo, Michigan, 1927.

Thorndike, E. L. The Teacher's Word Book. New York: Teachers College Publications, Columbia University, 1921.

Monographs

Alderman, Grover H. "Improving Comprehension Ability in Silent Reading," Eleventh Annual Conference on Educational Measurements, pp. 28-37. Bulletin of Extension Division, Vol. XI, No. 3. Bloomington, Indiana: Extension Division, University of Indiana, 1925.

Buswell, Guy Thomas. Fundamental Reading Habits: A Study of Their Development. Supplementary Educational Monographs, No. 21. Chicago: Department of Education, University of Chicago, 1922.

Carroll, Robert P. An Experimental Study of Comprehension in Reading. Teachers College Contributions to Education, No. 245. New York: Teachers College, Columbia University, 1926.

Gray, William S., with the cooperation of Kibbe, Delia; Lucas, Laura; and Miller, Lawrence W. Remedial Cases in Reading: Their Diagnosis and Treatment. Supplementary Educational Monographs, No. 22. Chicago: Department of Education, University of Chicago, 1922.

Gray, William S. Summary of Investigations Relating to Reading. Supplementary Educational Monographs, No. 28. Chicago: Department of Education, University of Chicago, 1925. Also, Supplements 1-5.

Judd, Charles H., and Others. Reading: Its Nature and Development. Supplementary Educational Monographs, No. 10. Chicago: Department of Education, University of Chicago, 1918.

Judd, Charles H., and Buswell, Guy Thomas. Silent Reading: A Study of the Various Types. Supplementary Educational Monographs, No. 23. Chicago: Department of Education, University of Chicago, 1922.

Kimmel, William G. Testing Pupil Progress in Community Life English. Supplementary Educational Monographs, No. 26. Chicago: Department of Education, University of Chicago, 1925.

Noonan, Margaret E. Influence of the Summer Vacation on the Abilities of Fifth and Sixth Grade Children. Teachers College Contributions to Education, No. 204. New York: Teachers College, Columbia University, 1925.

Tressler, J. C. "Silent Reading in the High School," in Contributions to Education, Vol. II, pp. 97-112. Edited by J. Carleton Bell and Ambrose L. Suhrie. Yonkers-on-Hudson, New York: World Book Company, 1928.

Unpublished Material

McCallister, J. M. "An Analytical Study of Reading Deficiencies of Junior High School Pupils." Unpublished Doctor's Dissertation, Department of Education, University of Chicago, 1929.

Morrison, Arthur F. "The Improvement of Instruction in Reading through Diagnostic and Remedial Measures." Unpublished Master's Thesis, Department of Education, University of Chicago, 1924.

"Remedial Cases in Reading." Unpublished Report of the Research Worker in Reading in the University High School, 1927-28 and 1928-29. Filed in the Office of the Research Secretary of the Laboratory Schools, University of Chicago.

"Remedial Reading." Unpublished Report of the Remedial Workers in Reading in the University High School, 1929-30. Secondary Work Reports, Vol. XI, Part 3. Filed in the Office of the Research Secretary of the Laboratory Schools, University of Chicago.

"Remedial and Corrective Reading." Unpublished Report of the Research Workers in Reading in the University High School, 1930-31. Secondary Work Reports, Vol. XII. Filed in the Office of the Research Secretary of the Laboratory Schools, University of Chicago.

Standardized Tests

- Burgess, Mae Ayres. A Scale for Measuring Ability in Silent Reading. New York: Russell Sage Foundation, 1921.
- Chapman, J. C. Chapman-Cook Speed of Reading Test. Philadelphia: J. B. Lippincott Company, 1923.
- Gray, William Scott. Silent Reading Test. Unpublished.
- Gray, William Scott. Standardized Oral Reading Paragraphs. Bloomington, Illinois: Public School Publishing Company.
- Inglis, Alexander. Test of English Vocabulary. Boston: Ginn and Company, 1923.
- Kelley, Truman L.; Ruch, Giles M.; and Terman, Lewis M. New Stanford Achievement Test. Yonkers-on-Hudson: World Book Company, 1931.
- Monroe, Walter S. Standardized Silent Reading Test III. Bloomington, Illinois: Public School Publishing Company, 1920.
- Otis, Arthur S. Self-Administering Tests of Mental Ability. Yonkers-on-Hudson: World Book Company, 1922.
- Sangren, Paul V., and Woody, Clifford. Reading Test. Yonkers-on-Hudson: World Book Company, 1928.
- Stone, Clarence R., and Murphy, Lucile C. The Stone Series of Narrative Reading Tests. Bloomington, Illinois: Public School Publishing Company, 1922.
- Thorndike, Edward L., and McCall, W. A. Thorndike-McCall Reading Scales. New York: Teachers College, Columbia University, 1920.
- Thorndike, Edward L. Visual Vocabulary Test. New York: Teachers College, Columbia University, 1916.
- Thorndike, Edward L. Test of Word Knowledge. New York: Teachers College, Columbia University, 1921.
- Whipple, Guy N. High School and College Reading Test. Bloomington, Illinois: Public School Publishing Company.

APPENDIX

The Appendix contains Form 1 of the battery of reading tests devised in this study. All four forms of the test battery and the test scores that comprised the original data used in the computations are contained in a complete unpublished form of this dissertation which is filed in the Library of the University of Chicago.

A SILENT-READING TEST FOR HIGH-SCHOOL PUPILS
(Grades 7-9)

Form 1

Do not write on the test, or turn a page, until you
are asked to do so by the examiner.

Name _____ Age _____
 last first yrs. mos.

School _____ Grade _____

Date _____ Boy _____ Girl _____
 (Place X on proper blank)

(After you have filled the blanks look at the directions.
The examiner will read them aloud and you will read them
silently.)

GENERAL DIRECTIONS.- This is a test to find out how well
you can read. It has three parts. Complete directions for doing
the test are given at the beginning of each part. Read all direc-
tions carefully and follow them exactly.

DIRECTIONS FOR PART I.- The first part of the test contains
a story for you to read and some questions for you to answer after
you have read the story. When the signal is given (but not before)
turn to page 2. Most of that page contains some questions printed
upside down. Do not try to read the questions, but look at the
bottom of the page where the story begins. The examiner will read
those lines aloud and you will read them silently. When the exami-
ner reaches the end of the last line, he will stop reading and you
will turn to the next page and continue to read. From time to
time, the examiner will say "mark." Put a circle around the word
you are reading when that signal is given, and keep right on read-
ing. When you finish reading, turn back to page 2, turn the page
around, and answer the questions which are based on the exercise.
Do not read any part of the material more than once. Read as fast
as you can read understandingly, but no faster, as you cannot
answer the questions unless you know what you have read.

Wait for the signal to turn to page 2.

When everyone is ready, the examiner will say:
"TURN TO PAGE 2 AND LOOK AT THE BOTTOM OF THE PAGE."

Scores will be recorded below

R _____

C1 _____

V _____

C2 _____

CT _____

(Add C1 and C2, multiply by 2,
and record after CT)

CS _____
(composite score)

These are the questions on the reading material. After you have read the exercise, you will turn back to this page, turn it around so that the questions are right side up, and answer them. Do not read the questions before you read the exercise.

Directions.- Underline the word or phrase which should be used to complete each statement, and place its number in the parentheses at the end of the line. Do not turn back to the reading exercise.

Example.- This is a test of (1) arithmetic (2) spelling (3) reading (4) geography (5) history (3)

1. The cub whimpered because (1) he was cold (2) he was tired (3) he was hungry (4) he was lost (5) he was afraid ()

2. The first kill the bear made was (1) a pair of ruffed grouse (2) a pair of ptarmigan (3) a white rabbit (4) a young deer (5) a musk ox ()

3. When the bear caught sight of the musk-oxen she (1) stopped and watched them (2) rushed at them at once (3) approached them cautiously (4) tried to creep up to them (5) made a wide detour around them ()

4. As soon as they saw the bear, the little bulls (1) swung to meet her (2) rushed at her (3) scattered in all directions (4) set up a loud bellowing (5) deserted the cows and calves ()

5. The bear retreated from the musk-oxen as if in terror because she (1) was afraid of the bulls (2) wanted them to think she had gone away (3) was anxious for the safety of her cub (4) hoped to be pursued (5) was in a playful mood ()

6. After the bear left the musk-ox herd, she moved (1) out upon the open plain (2) toward a range of mountains (3) along a river (4) down a narrow gorge (5) toward the open sea ()

7. When she heard the tinkle of breaking ice the bear (1) couched low (2) ran toward the sound (3) looked all about her (4) snuffed the air (5) pushed the cub to safety ()

8. As the bear crept upon the seal the cub (1) followed behind her (2) stood at a distance and watched (3) crouched behind some ice (4) whimpered in anxiety (5) began to howl ()

9. The seal was prepared for an attack from (1) any side (2) the land (3) the open sea (4) any direction except the south (5) the north and east ()

10. When the seal leaped into the water, the bear (1) was obliged to loosen her grip (2) sprang in after him (3) found the odds too much for her (4) sank her teeth deeper into the seal (5) swam toward him ()

(When you have finished the foregoing questions, close the booklet and wait for further instructions.)

QUESTIONS ON FORM 1

NANOOK AND HER CUB*

It was winter in the far north. A wasteland of rocks and ice and snow stretched for many miles in every direction.

At the zenith of a black-violet sky, the Great Bear prowled around the Pole Star, guarded by a frost-white moon. On the earth below, another great bear wandered restlessly through the night in quest of something to eat. This unending search for food was the price that Nanook, the she-bear of the Arctic paid for life. The half-grown cub who followed her, whimpered hungrily in his throat, for it was long since he had tasted food.

At his cry, the grim mother checked her stride; but even as she stopped her nostrils caught the faint, unmistakable scent of warm flesh and blood. Swinging her long, pointed head to one side, she moved silently, on paws padded with thick white hair, toward a hummock of new-fallen snow. Just as she reached it, the little hillock exploded like a bomb and out whirled a pair of ptarmigan, birds like the ruffed grouse of lower latitudes, but white as the snow in which they had hidden. Before they had freed themselves from the snow the great ice-bear struck them down. A moment later a cloud of feathers was all that was left of them, and the bears wandered on through a country desolate as the surface of the burnt-out moon.

Hours later, as the two reached a plateau high above the confusion of ice and rock, there came to the straining nostrils of the leader a sweet, warm reek as of cattle, and scattered black figures showed dimly against the snow. At the sight and scent, the huge bear shot forward like a sprinter from his mark. As she charged, a score of sturdy, wide-horned little bulls swung to meet her, with the swift discipline of a musk-ox herd. They had short legs, hair so long that it trailed on the snow, and all the swift alertness of mountain sheep. Behind the bulls was an inner circle of cows, within which were a dozen frightened calves. So swiftly had the trained herd acted, that by the time the bear reached them, she was confronted by a circle of sharp curved horns.

Hungry as she was, with a whimpering cub at her flank, the gaunt bear stopped—as well she might, for not even the grim white wolf, the craftiest of arctic hunters, can penetrate the defence of a musk-ox herd. Around and around the bristling hedge of horns she paced, but there was never a break in the threatening circle. With little grunts of defiance, the veteran fighters of the herd, who had faced danger many times before, heartened the younger cows and the trembling calves behind them, and everywhere their front remained unbroken. Once or twice the bear rushed forward, as if intending to charge in spite of all consequences, on the chance that some over-enthusiastic bull would leave the safety of the circle and meet her half-way. Again turning her back, she retreated slowly as if in sudden terror, hoping to be pursued, but the herd paid no attention to any such simple stratagems. Only a line of lowered glaring eyes and a ring of fatal horns confronted her until at last, with many a backward glance, she lurched away, followed by her cub, in search of easier prey.

Disappointed by the musk-oxen, the bear peered into every nook and cranny as she moved across the frozen wilderness hunting for food. Her course led her at last through a maze of basalt boulders, ice-blocks and pans toward a stretch of open sea. As she approached that lonely water under the orange glare of the Northern Lights, her keen ears caught the tiny tinkle of breaking ice where the black water close to the shore had begun to skim over in the cold. At the sound, the bear couched low, with the cub imitating

her every movement. In the dim light, a long dark shape floundered awkwardly up on the shore. There, poised where he could see all about him and shoot down into the water at the first sign of danger, lay a huge harp seal—the "saddleback" of the sealers. This was an old bull who weighed all of 800 pounds. He had a dingy white body blotched with black, and was far larger than the spotted leopard seal, the ringed seal or even the gray-bearded seal.

Without a sound, the white bear disappeared in the dark like a ghost, while her cub crouched behind a hillock of ice. Making a long detour, the fierce huntress reached the shore at a point hidden by heaped masses of ice from the sight of the seal and, slipping soundlessly into the water, swam far out, only her black muzzle showing above the surface. When at last she reached a point in the bay opposite the place where the saddleback was, she swam toward the shore, silent as a shadow.

The great seal lay with his back to the water, continually testing the air with his sensitive nostrils and ready at first sight, sound, or scent of danger to rush into the water. Wary as he was, he suspected no danger from the open sea which he had just left. Suddenly, his keen ears caught the sound of dripping water and a little grating noise at the edge of the shore, which might have been made by floating ice. However, the motto of the wise seal is: "Make sure!" As for the unwise, they are dead, victims of their own imprudence.

Wherefore, this one, at the first slight noise behind him, swung around, only to see a vast shape creeping toward him. At the sight, the seal hurled his mottled bulk toward the open water. Swift as he was, the white death below was swifter. Even as the saddleback reached the water, the fierce muzzle of the bear shot forward like a striking snake and gripped the seal's round head. Despite the bear's weight the rush of the saddleback carried her down into the black water. Few land-animals indeed would dare the depths with such a swimmer. Maddened by hunger, however, the bear paid no attention to the odds of the water that were against her, but sank her teeth deeper and deeper through the seal's thick pelt. A moment later, pushing the lifeless body of her prey before her, she swam slowly up through ten feet of black water laced and lit by the phosphorescence of Arctic seas. Then she landed and without an effort swung the body of the seal out upon the rock, and hissed sharply through the still air. At the sound her hungry cub hurried from his hiding place so fast that he rolled over and over down the slope like a ball, nor ever stopped until he joined her in a feast of rich, pure, life-giving seal meat.

When at last they were both full-fed, the old bear dragged the carcass along the shore and hid it in a cave near the top of an ice-cliff which jutted out over the water. There, curled up together, the two slept, plunged in that snug content which food and warmth bring.

(Now turn back to page 2 and answer the questions.)

*This story is adapted from "Nanook" by Samuel Scoville, Jr., of Philadelphia, and is used with the author's permission.

Part II

A TEST OF READING VOCABULARY

Form 1

DIRECTIONS.- In each of the exercises below you are to select that one of the five words or phrases which most nearly corresponds in meaning to the word which is underlined in the sample sentence or expression. Draw a line under the word thus selected and place its number in the parenthesis at the right. Notice the example below:

EXAMPLE.- Let us slacken the pace.

(1)quicken (2)change (3)reduce (4)continue (5)hold (3)

1. They adapt themselves well.

(1)please (2)adjust (3)carry (4)conduct (5)consider ()

2. An aggressive person.

(1)agreeable (2)handsome (3)large (4)vigorous (5)excellent()

3. The apparatus is expensive.

(1)appliance (2)clothing (3)reading-matter (4)engine (5)experiment ()

4. We aspire to rule.

(1)refuse (2)fear (3)desire (4)ask (5)appear ()

5. They barter goods.

(1)buy (2)manufacture (3)loan (4)exchange (5)own ()

6. Betake to the country.

(1)hurry (2)go (3)wander (4)send (5)return ()

7. The government is in chaos.

(1)session (2)order (3)confusion (4)excellent-condition (5)defeat ()

8. The commendation is deserved.

(1)success (2)blow (3)popularity (4)good-fortune (5)approbation ()

9. His action received condemnation.

(1)approval (2)applause (3)censure (4)sympathy (5)contempt ()

10. They conspire against the ruler.

(1)plot (2)fight (3)move (4)speak-openly (5)write ()

11. Defer the payment.

(1)make (2)postpone (3)request (4)discharge (5)record ()

12. We depute you to act.

(1)forbid (2)ask (3)authorize (4)entreat (5)require ()

13. The diplomat is well-known.

(1)musician (2)artist (3)tactful-person (4)leader (5)breakfast-food ()

14. It is easy to dupe him.

(1)lose (2)forget (3)deceive (4)win (5)defeat ()

15. Boilers emit steam.
(1) eject (2) contain (3) make (4) retain (5) use ()
16. A sad epitaph.
(1) reply (2) verse (3) sermon (4) message (5) inscription ()
17. We exhort you to go.
(1) urge (2) wish (3) forbid (4) expect (5) hire ()
18. A fervent prayer.
(1) long (2) intelligent (3) loud (4) hypocritical (5) eager ()
19. He went into a frenzy.
(1) inn (2) fright (3) trance (4) fury (5) carriage ()
20. You should gauge its speed.
(1) know (2) estimate (3) observe (4) match (5) retard ()
21. He was defeated by guile.
(1) deception (2) overwhelming-strength (3) jealousy (4) hatred
(5) enemies ()
22. A hitch in the plans.
(1) error (2) surprise (3) disagreement (4) construction (5) reorganization ()
23. Impenetrable gloom.
(1) faint (2) star-lit (3) fearful (4) hostile (5) dense ()
24. The innovation proved successful.
(1) invention (2) change (3) publication (4) attempt (5) operation ()
25. An intricate passage.
(1) deep (2) dark (3) interesting (4) complicated (5) narrow ()
26. The legacy is large.
(1) fortune (2) request (3) estate (4) stadium (5) bequest ()
27. A loathsome scene.
(1) beautiful (2) quiet (3) fearful (4) peaceful (5) disgusting ()
28. A meager salary.
(1) large (2) desirable (3) scanty (4) monthly (5) adequate ()
29. A minor position.
(1) important (2) inferior (3) dangerous (4) insecure (5) influential ()
30. A motley crew.
(1) well-dressed (2) mixed (3) rugged (4) trustworthy (5) brave ()
31. We can muster a thousand.
(1) assemble (2) defeat (3) see (4) manage (5) feed ()
32. A notch in the wood.
(1) knot (2) stain (3) nail (4) crack (5) nick ()
33. His plaint was heard.
(1) voice (2) musical-instrument (3) oration (4) howl (5) complaint ()

34. A precise statement.

- (1)angry (2)forceful (3>false (4)accurate (5)truthful ()

35. The reaper is here.

- (1)mower (2)planter (3)farmer (4)harvester (5)tractor ()

36. Replenish the supply.

- (1)repair (2)use (3)ask for (4)measure (5)save ()

37. A deep revery.

- (1)canon (2)stream (3)dream (4)speech (5)secret ()

38. A salutary suggestion.

- (1)serious (2)humorous (3)ridiculous (4)happy (5)healthful ()

39. The roe darted away.

- (1)rabbit (2)quail (3)boat (4)mountain-goat (5)deer ()

40. The boat was on the shoal.

- (1)surface (2)sunken-log (3)sea (4)bottom (5)shallow-place ()

41. The liquid will solidify soon.

- (1)explode (2)boil (3)harden (4)cool (5)evaporate ()

42. Stifle the suggestion.

- (1)publish (2)observe (3)consider (4)suppress (5)forget ()

43. They wish to subvert the law.

- (1)obey (2)destroy (3)pass (4)avoid (5)break ()

44. The supremacy of arms.

- (1)superiority (2)clash (3)strength (4)brilliance (5)use ()

45. The beauty of the tabernacle.

- (1)mountain (2)parsonage (3)temple (4)monastery (5)painting ()

46. The trend is to the west.

- (1)course (2)view (3)landscape (4)road (5)village ()

47. Usury is forbidden.

- (1)drunkenness (2)cheating (3)selling-liquor (4)murder
(5)unlawful-interest ()

48. Our eyes beheld much verdure.

- (1)filth (2)grandeur (3)vegetation (4)blood (5)waste-land ()

49. Wend northward.

- (1)face (2)go (3)run (4)walk (5)look ()

50. They wrangle constantly.

- (1)laugh (2)fight (3)quarrel (4)play (5)cry ()

(Go over this part of the test again to see if you have done as well as you can with it. Then close the booklet and wait for instructions to start the last part.)

Part III

A TEST OF COMPREHENSION IN SILENT READING

DIRECTIONS

This is an exercise to find out how well you can read and get the meaning. It is made up of several paragraphs, followed by questions. Read each paragraph and answer the questions based on it. You may read a paragraph again if necessary. Make your answers short. Usually one or two words will be enough. Where several choices are given in a question, underline the one you think is correct. Work carefully, but do not waste time. You will be given 20 minutes. Do the sample exercise below.

Sample Exercise

It was a cold, stormy, December day. A blinding snow swept down upon the city from the north. A stream of afternoon shoppers hurried along the street with heads bent against the wintry blast.

1. What season of the year was it? _____
2. Was the sun shining? _____

Underline the word which tells how the shoppers walked.
(a)slowly (b)leisurely (c)gayly (d)rapidly (e)proudly

If you have questions, ask them now, as no questions will be answered after the exercise starts.

Wait for the signal to start.

To be read by the examiner when all are ready:

TURN TO THE FIRST PARAGRAPH AND BEGIN READING

Read the paragraph and answer the questions. Read it again if you need to. If several choices are given in a question, underline the one you think is correct.

Shaggy was captured when he was a tiny cub and was carried to a village where he was sold to a young army officer who was very fond of animals. Here he was kindly treated and he soon grew to like his life in captivity. The young bear's master gave him a warm den and plenty of food, and often took him for a walk.

1. What kind of animal was Shaggy? _____
2. What was his master's feeling toward animals? _____

-
3. How do you think Shaggy felt in captivity?
(a)afraid (b)contented (c)affectionate (d)lonely (e)angry

------(Do the next paragraph)-----

The penguin is a very curious bird and takes great pains to inspect any strange object he may see. Once when some Antarctic explorers were waiting for the ship to take them home they lived in little tents which they had pitched on the snow about fifty yards from the edge of the sea. Parties of penguins from the neighboring rookery often landed there, and almost invariably the first thing they did on seeing the tents was to go up the slope and inspect them. They would walk all around the tents and would often stay to doze by them for hours. Some of them, indeed, seemed to enjoy the companionship of the men, and none of them showed any sign of fear.

1. In what part of the world did these penguins live?
(a)near the South Pole (b)in North America (c)near the equator
(d)in Iceland (e)near the North Pole.
2. What did the penguins do when they first saw the tents?

-
3. This selection indicates that the attitude of penguins toward man is one of (a)fear (b)suspicion (c)indifference
(d)inquisitiveness (e)great friendliness.

------(Do the next paragraph)-----

Drowned valleys are the lower parts of stream valleys which have sunk below sea level. Many rivers are in this condition. The sea extends up the Delaware River to Trenton, and Chesapeake Bay with its many arms is the drowned valleys of the Susquehanna and its former tributaries. The lower St. Lawrence is perhaps the greatest example of a drowned valley in the world. Some of the world's most famous harbors, as San Francisco Bay, Puget Sound, the Scottish firths, and the estuaries of the Thames and the Mersey are drowned valleys.

1. Write the phrase which tells the altitude of drowned valleys. _____

2. What is perhaps the greatest example of a drowned valley in the world? _____

3. Name a bay, not mentioned as one of the most famous harbors in the world, which is said by this writer to be a drowned valley. _____

------(Do the next paragraph)-----

The words "public duty" do not necessarily mean official duty, although they may include that. They mean simply that constant and active practical participation in the details of politics without which, upon the part of the most intelligent citizens, the conduct of public affairs falls under the control of selfish and ignorant, or crafty and venal men. They mean that personal attention—which, as it must be incessant, is often wearisome and even repulsive—to the details of politics, attendance at meetings, service upon committees, care and trouble and expense of many kinds, patient endurance of rebuffs, chagrins, ridicules, disappointments, defeats—in a word, all those duties and services which, when selfishly and meanly performed, stigmatize a man as a mere politician; but whose constant, honorable, intelligent, and vigilant performance is the gradual building, stone by stone and layer by layer, of that great temple of self-restrained liberty which all generous souls mean that our government should be.

1. Write the four words which describe the kinds of men who will control public affairs if intelligent citizens do not participate in politics actively and positively. _____

2. Personal attention to the conduct of public affairs must be (a)continuous (b)selfish (c)repulsive (d)troublesome (e)disappointing.

3. What do all generous souls mean that our government should be? _____

4. Does this paragraph lead one to believe that intelligent citizens will find the work of politics easy and pleasant? _____

------(Do the next paragraph)-----

An interesting issue was settled not long ago at the preparatory disarmament conference at Geneva. This question was whether in fixing the ratios of personnel among the navies the plan should be global or by categories. This means shall the ratio of men to be permitted each navy refer merely to the whole number of men in each service or shall the proportion of commissioned officers, noncommissioned officers and enlisted men within the whole number also be fixed. The British were in favor of the first method because if they could maintain a very high percentage of officers, their large merchant marine could be supplied promptly with a sufficient number of trained officers in case war should break out. The French, supported by the soviets, contended for the fixing of categories. America took the British side of the question, and this plan finally prevailed.

1. If the proportion of commissioned officers, noncommissioned officers and enlisted men permitted each navy is fixed, the

plan is _____

2. What delegates were in favor of the first method?

3. Write the name of the plan which prevailed _____

4. The British stand on the question was influenced by the fact that they (a)feared the French (b)had too many officers (c)desired many officers (d)needed more enlisted men (e)wanted to protect small nations.

------(Do the next paragraph)-----

Esthetics is the philosophical study of the human experiences connected with the fine arts, with beauty, and with taste in all relationships. Its purpose is to bring order and understanding into the multitude of bewildering phenomena covered by the word "beautiful" taken in the widest sense. Its pursuit requires that rare combination of keen sensibility and clear thinking, cultivated taste and logical rigor. Oftentimes one of these abilities outstrips the other, and mediocre or bad theories are then likely to result. The esthetician usually starts from personal experiences that are abundant and ever renewed and seeks, by critical analysis and interpretation, to establish generalizations that are illuminating and conducive to a fuller enjoyment of art.

1. Esthetics deals with what phases of human experience?

2. Esthetics seeks to bring order and understanding into what phenomena?

3. According to this paragraph, what kinds of thinking does the esthetician do in order to establish generalizations that are illuminating and conducive to a fuller enjoyment of art?

(Look over this part of the test again. Perhaps you can improve upon some of your answers. When you are sure that you have done as well as you can with the test, close the booklet and wait quietly for the others to finish. Read or study if you have a book with you.)

